THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse III - Paul Sabatier*
Discipline ou spécialité : *Ecologie, Biodiversité, Evolution*

Présentée et soutenue par *Florent BLED*
Le *18 mai 2010*

Titre : *Tests d'hypothèses en dynamique des populations fragmentées: développement et applications de modèles d'occupation des sites*

Title: *Tests of hypotheses in fragmentated population dynamics: development and applications of site occupancy models*

JURY

Sergine Ponsard
Roger Pradel
Michael C. Runge
Peter B. Pearman
Emmanuelle Cam
Jérôme Dupuis

Ecole doctorale : *SEVAB*
Unité de recherche : *UMR 5174 Evolution et Diversité Biologique*
Directeur(s) de Thèse : *Emmanuelle Cam et Jérôme Dupuis*
Rapporteurs :
Roger Pradel
Michael C. Runge

In memory of Max Bled

To my goddaughter, Charlotte

Abstract – The knowledge of distribution and local abundance of organisms is a major element of population studies. It has several implications in areas ranging from basic research in population dynamics to more applied topics such as conservation and population management. The distribution and abundance of organisms involves both physical and biotic processes that vary spatially and temporally, and additionally, are typically non-linear and dynamic. While a common type of data used to assess species occurrence is binary presence-absence data, classical approaches to the development of spatial models for binary processes (i.e. "occupancy processes") present three important deficiencies.

First, they do not explicitly accommodate sampling uncertainty in the form of false absences. Hierarchical modeling accounting for uncertainty in detection permits consideration of this problem, but even in that case, there are problems in accounting for sampling or in conditioning model on the presence of the species of interest in the area where the sampling protocol is conducted. Secondly, there is an lack of spatio-temporal models of occupancy, especially in the framework of hierarchical modeling. Finally, most of existing models are phenomenological models and do not explicitly consider underlying ecological mechanisms. However, it is obvious that when it is possible to use ecological mechanisms to describe and predict site occupancy, this is an advantage.

In this thesis, I have developed spatio-temporal occupancy models for dynamical ecological processes in order to respond to these limitations of actual site occupancy modeling. Such models are critical for modeling the spread of important diseases by wildlife populations, the spread of invasive species, and range dynamics in the presence of a changing environment. A general hierarchical modeling framework is proposed for spatio-temporal dynamic occupancy processes in the presence of uncertain observations. In addition to incorporation of sampling uncertainty in the observations, a key component of this research is related to the incorporation of scientific knowledge in the model. One of the primary goals of my research has been to develop scientifically-meaningful and statistically rigorous models for spatio-temporal dynamic occupancy processes in the presence of uncertainty that can be used at several different geographical scales. Moreover, I wanted to propose a structure of models adaptable to several topics, correcting issues that appeared in previous general occupancy models.

Simultaneously with modeling issues, three main ecological topics are addressed in this thesis. The first is related to invasive species that are commonly claimed as the second threat

on biodiversity. In order to apply a relevant response to the potential danger associated with invasions, it is essential to understand invasion mechanisms and dynamics. Focusing on the propagation of an invasive species in USA (the Eurasian collared dove), I have developed a general hierarchical spatio-temporal dynamic occupancy framework accommodating the probabilistic automata case and generalizing it to the realistic situation in which there is uncertainty in the observations. Secondly, I have focused on nesting sites in the kittiwake. Here, I have proposed a model that explicitly encompasses hypotheses developed in the framework of habitat selection to estimate nesting site occupancy dynamics within a cliff. Evolutionary ecology provides a conceptual framework to address the relationship between individual decisions and habitat features. We estimated demographic processes of site persistence (an occupied site stays occupied), and colonization through two subprocesses: first colonization (site creation) and recolonization (a site is colonized again after being deserted). Moreover, my model incorporated local and neighboring breeding success and conspecific density in the neighborhood. Finally, I have applied the work addressing the issue of conditioning on the presence of the species in the area of interest and its implication for estimation of occupancy rate. I have estimated the evolution of the occupancy rates of several bird species in a French forest regarding of climatic changes and considering spatial heterogeneity. Our model is used to study the impact of three consecutive particularly cold winters on a selected set of bird species. I have focused on a limited range of factors that might influence the response of some bird species to climatic changes (sedentary vs migrating species ; biogeographical origins).

Conservation efforts require the ability to understand range and occupancy dynamics as a function of changes in dynamic features such as land use and climate. Such understanding will permit prediction of occupancy changes that are likely to accompany future changes and hopefully will permit informed attempts to mediate changes in occupancy that are viewed as undesirable. I have discussed in the last part of this thesis the future ways site occupancy modeling could take and that will ultimately lead to the understanding of occupancy dynamics required for conservation and management.

Acknowledgements

First of all, I would like thank EDB laboratory and the USGS Patuxent Wildlife Research center where I have conducted my research. I would also like to thank the international relationships service of the Université of Toulouse 3 and the "Conseil Général du Finistère" for their funding that made possible my trips to the USA.

I would like to thank Brigitte Crouau-Roy and the whole team of researchers of the EDB laboratory for letting me works with them.

I deeply express my gratitude to Pr. Emmanuelle Cam and Jérôme Dupuis who have been the directors of this PhD for all their support, help and scientific emulation they have provided to me, probably not even realizing how much it was a great experience to work with them. During more than three years, I have learned from them much more than just scientific knowledge.

I would particularly like to thank Sergine Ponsard, Michael C. Runge, Roger Pradel and Peter B. Pearman who have accepted to participate in my PhD jury. I am grateful to the reviewers for their helpful comments.

I would also like to thank Sovan Lek and Claude Maranges who are the directors of the Doctorate school SEVAB for their help in the submission process of this thesis. I am also grateful to Dominique Galy, Dominique Pantalacci, Peggy Leroy and Marie-Martine Bègue who have been of great help in the administrative tasks that I had to experience during these three years.

I am grateful to Andy Royle for accepting me in the USA, for his help both scientifically and personally once I have been there. I also thank him for all the interesting conversations we had, and for granting me access to BBS data. I would also like to thank Bill Link, Jim Hines and Jim Nichols for their comments on my work when I was over there.

I thank Marc Kéry, Béni Schmidt and Robert Dorazio for letting me attend the workshop on hierarchical models in ecology in Zurich.

I am grateful to Jean-Yves Monnat for accepting me on fieldwork, for having been in my PhD committee and for the experiences I had thanks to him on Cap Sizun cliffs, handling

kittiwakes' chicks. I am also grateful for sharing his vast knowledge, not only on the kittiwakes, but also as a great naturalist, and cook.
I thank Jean Joachim for our interesting exchanges and for letting me know more about the bird community in South France, and for letting me use his data.
I wish to thank all the people who have at one time or the other participated to the kittiwake survey and to the breeding bird survey that permit this work.

I would also like to thank Etienne Danchin and Philippe Heeb for the comments they gave me during this thesis.

I am grateful to Pierrick Blanchard, Pascal Mirleau and Jean Baptiste Ferdy for having been such great office colleagues. Pierrick.... Keep on doing GRS, you'll be the next Paul Hunth!

I am thanking Jean-Sébastien Pierre who has accepted to participate in my PhD committee, and Denis Poinsot. Both of them opened a new area of interest in Science for me. By their passions, their open-mind and their abilities to share all of these, they have significantly ($p<0.001$) marked my journey.

I am deeply grateful to Karl Doré, Laure Simonin, Brice Maret, Laurent Kernéis, Malvina Rognant-Lak and Charlotte Veyssière. You have been here during the best and the most difficult times and I can not thank you enough for this.

I am thanking Lise Aubry, David Koons, Elise Zipkin, Sarah Converse, Beth Gardner, Julien Martin, Adam Green and Suzie Ponce for their support, and for all the good moments we had while in the USA.

Thanks to all the young generation of the EDB community for all the fun and activity they have created in the 4R3 building and in my life: Pierre-Jean Malé, Juliane Casquet, Kyle Dexter, Candida Shinn, Laetitia Buisson, Clément Tisseuil, Frank Jabot, Erwan Quéméré, François Chassagne. I am also thanking Ael Kermarec, Elodie Stéphan, Eric Bard, Sébastien Goumon, Masayoshi Sugimoto, Camille Ansquer, Lyndsay Aitken, Mathieu Rognant, Aurèle, Gilles, Cécile, Delphine Gallou-Papin and Raphael Papin-Gallou, Adrien Chorein, Jean Calice, Marine El Adouzi, Vincent Cuminetti, Cléx, Marine, Mik, the A.D.L.M. who have been part of my life, and helped me to be what I am now. And thanks again for all of you termajis that help make this life better! I also deeply thank Speed Burger and la Boite à Pizza for their unconditional help in the healthy nutrition process that occurred during this PhD.

Finally, I would like to thank from all my heart my family, especially my mother, my godmother, my godfather (and the 2kg Gilt-head bream), and my aunt Rachel. If you weren't

here, this thesis would not have been possible. You taught me how to fish instead of simply giving me a 'tacot'. Thank you for this. I also thank my brother for everything we share, and I'd like you to know how much I'm proud of you. I'll never forget all of you.
On a fluffier note, thanks to Radjah that help the life to be bouncier!

I'd like to end these acknowledgements by dedicating this thesis to my father and my goddaughter. Charlotte, your smile and your incredible blue eyes are the sunrays that permit to see a blue sky behind the clouds, even in Brittany. You are my sunshine and I could not love you more, but I keep trying to do so for all my life. Dad, without you, I don't know where I would have been. You are a model for me and I'm proud to be your son. This thesis is yours.
Love you...

SUMMARY

List of tables

List of figures

CHAPTER I

INTRODUCTION

1

Ecology has been defined as "the scientific study of the interactions that determine the distribution and abundance of organisms" (Krebs, 1978). The knowledge of distribution and local abundance of organisms is a major element of population studies. This is the corner stone in areas ranging from basic research in population dynamics (Krebs, 1978) to more applied topics such as conservation and population management (Kendall, 2001 ; MacKenzie *et al.*, 2005 ; Williams *et al*., 2002 ; Nichols, 2004). As a result of human activities (e.g. agriculture, deforestation, dam or barrage, development of highway systems), fragmentation of environment in which individuals and species live has dramatically increased (Vitousek *et al.*, 1997). Individuals of mobile species often have to live in a set of subdivisions (or sites/patches) of their natural habitat, and therefore may have to move from patch to patch.

It is widely recognized that many useful models of biological systems can be developed from binary "presence/absence" data. Such data are relatively inexpensive and easy to obtain, and may yield a maximally informative reduction of data collected under loosely organized, complex or varying protocols. Indeed such data have a long history of use in ecology, beginning with the investigation of spatial patterns of association between two species (Forbes, 1907 ; Dice, 1945), moving to the development of theory of species distribution patterns (MacArthur & Wilson, 1963, 1967 ; MacArthur, 1972), including hypotheses about community assembly rules (Diamond, 1975) and nested subset community structures (Patterson & Atmar, 1986 ; Patterson, 1987). These data can even be used to estimate species abundance (Royle & Nichols, 2003 ; Royle *et al.*, 2005 ; Pearce & Boyce, 2006). The last two decades have seen a resurgence of interest in macroecology (Brown & Maurer, 1989 ; Brown, 1995 ; Rosenzweig, 1995 ; Gaston & Blackburn, 2000 ; Bell, 2001 ; Hubbell, 2001), involving the study of distribution and abundance or organisms at large spatial and temporal scales. The fields of biogeography and landscape ecology have also recently emerged as important subdisciplines in ecology, and share many of the same goals as macroecology.

Conservation-oriented studies of spatio-temporal dynamics are especially timely. For example, the modeling of species distribution dynamics will be useful in developing predictions about distributional changes expected to accompany both land use changes and active land management. Natural range expansions by a species into areas inhabited by an “inferior” competitor species may result in range contractions of the competitor (e.g., barred owl expansion into northern spotted owl range ; Anthony *et al.*, 2006), possibly deserving management actions. Investigations of species range dynamics over the last several decades can provide a basis for predictions about future range changes in response to global climate

change. Invasive species have become a major problem throughout the world. Investigations of the spatio-temporal dynamics of invasive species will permit predictions about future spread as well as about the likely effectiveness of management actions designed to halt such spread (e.g. Wikle, 2003 ; MacKenzie *et al.*, 2005).

In spatial statistics, spatialization of data (e.g. "presence/absence") is commonly accounted for using four types of records. (i) The spatial structure can be reflected by surface recording: data are related to area delimited by borders:. This approach has long been used in socioeconomics (e.g. Geary, 1954). (ii) In the case of punctual records, data refer to two coordinates (x,y), and it is possible to associate surface to punctual records choosing a particular point by spatial unit. (iii) It is also possible to define the spatial structure by neighboring relationships. For example, two area units where records are available are considered as neighbors if they share a common border. (iv) Finally, the simplest case is the distance, where the spatial structure is accounted for by canonical Euclidian distance between points (Chessel & Thioulouse, 2003). In all these cases of spatialized data, the site or patch (i.e. points or areas where data are recorded) occupancy can be one of the variable of interest. As a matter of fact, the spatial pattern of plants and animals is an important characteristic of ecological communities. This is usually one of the first observations made in studying any community and is one of the most fundamental properties of any group of living organism (Connell, 1963). Considering a simple spatial structure, three basic types of patterns are recognized in communities: random, clumped and uniform (Ludwig & Reynolds, 1988). On the time scale, a patch or site is occupied or unoccupied; when an occupied patch is deserted between two units of time (e.g. years or seasons), it is an extinction event. Conversely, when a patch is occupied in successive sampling occasions, it is named persistence or “survival” by analogy with survival probability of individuals in populations. An unoccupied patch that becomes occupied corresponds to an event of colonization.

These processes of occupancy dynamics will occur at different time and spatial scales, whether we consider individuals, species, or even communities. When focusing on individuals, ecological systems are fundamentally systems that present a hierarchy based on spatial and time scale (cf. table 1, from Royle & Dorazio, 2008).

Table 1. Examples of ecological scales of organization. Each of these systems is characterized by a specific 'size' parameter, 'N', that is often the quantity of interest in static systems. In dynamic systems, there are analogs of survival and recruitment probabilities, which are usually described by extinction and colonization parameters in metapopulation and community systems. (Royle & Dorazio, 2008)

	Static system	Dynamic system
Population of individuals	N = population size	ϕ = survival γ = recruitment
Population of populations (metapopulation)	N(s) = populations size spatially indexed $\psi(s) = Pr(N(s)>0)$	$1 - \phi$ = extinction γ = colonization
Population of species (metacommunity)	N = species richness	ϕ , γ
Population of communities (metacommunity)	N(s), $\psi_i(s)$	ϕ , γ

These analogies are self-evident here and they stay true when we focus on sites. For example, depending on the spatial scale considered, we will always have a set of patches (e.g. ponds, forests, geographical regions, continents). The only change will be in the parameters of the considered ecological scale that will influence the site occupancy dynamics. At a small spatial range (e.g. breeding site choice within a colony by a seabird, migration of batrachians between ponds), it is a small ecological scale where site occupancy will result from individual behavior of habitat selection. At a larger scale, like a country or a geographical area, site occupancy may be influenced by population parameters (e.g. population size in each patch). Therefore, in these conditions, site occupancy modeling can be used to address a large range of apparently different biological topics. Relying on the same processes of colonization and persistence, , underlying ecological mechanisms change depending on the specific ecological scale. Moreover, site occupancy depends of physical and biotic processes that vary both spatially and temporally (Hanski & Gilpin, 1997) and consideration of these variations should be of prime interest when modeling site occupancy.

Classical models in ecology and epidemiology – At no period in the history of ecology has the spatial structure of populations and communities been entirely ignored, but the part that space plays in determining ecological patterns and in molding processes has been viewed

very differently across time (McIntosh, 1991). In the 1960s and 1970s, theoretical ecology was largely focused on issues other than spatial dynamics (May, 1976), with notable exceptions (Mac Arthur and Wilson, 1967), and field ecologists tended to follow suit (Hanski & Simberloff, 1997). Today, space is in the forefront and is introduced in various ways into all field of ecology and population biology more generally. Whether one is interested in processes occurring at the level of genes, individuals, populations, or communities, spatial structure is widely seen as a vital ingredient of more powerful theories, and good empirical work involving space is seen as a great challenge (Kareiva, 1990).

In epidemiology, when studying the spatial distribution of a disease, it is recognized that there are basic models which are usually assumed to apply, at least as a starting point, in the analysis of case event or count data. A key role is played by the Poisson process, related point process models and the Poisson distribution. When fundamental assumptions of these models are not completely met, more complex models (often consisting of random effects) must be invoked. (Lawson, 2006). The development of applications of models for spatial point processes has gone through various phases. Many early developments took place in ecological applications and, in particular, forest science (Matérn, 1986). In these applications, it was often the case that relatively large realizations of points were observed (e.g. plant communities or forests), mainly in a homogeneous environment. This led to the analysis of models for events in homogeneous environments, and to special methods for "sparsely sampled" problems, which are found particularly in ecological examples (Lawson, 2006). In these early studies a number of basic models for point processes were applied. Among these models the three most important in applications were complete spatial randomness (CSR), spatial cluster processes and spatial inhibition processes. Diggle (2003), Ripley (1981) and Cressie (1993) provide reviews of this work.

In the field of island biogeography, migrations rates among geographical units are modeled as functions of island size, distance to a mainland, and sizes of mainland and island population units (MacArthur & Wilson, 1967). Migration rates and sources of variation in these rates are relevant to modeling in population genetics (e.g. island *versus* stepping stone *versus* more general isolation-by distance models; Crow & Kimura, 1970). As is the case with many aspects of population-dynamic modeling, human demographers were the first to incorporate multiple locations into projection matrix models (Rogers, 1966, 1968, 1975, 1985, 1995 ; Le Bras, 1971 ; Schoen, 1988). These multiregional matrix models now are being applied in animal ecology (e.g. Fahrig & Merriam, 1985 ; Lebreton & Gonzalez-Davilla, 1993 ; Lebreton, 1996 ; Lebreton *et al.*, 2000).

On the basis of population dynamics parameters such as local population density, population growth rate and dispersal distance, several mathematical formulations have been developed to characterize and model the rate of spatial advance of spreading populations. Examples include reaction–diffusion equations and their extensions (Fisher, 1937; Skellman, 1951; Ortega-Cejas *et al.*, 2004) and integrodifferential equations (Van den Bosh *et al.*, 1990, 1992; Kot, 1992; Neubert & Caswell, 2000). In the case of spatial disease modeling, there is a wide variety of models variants available in the space-time extension. For example, semi parametric models may be useful and spatial spline models are easily extended to spatiotemporal situations. Autologistic models (models in which the modeling of spatially-distributed continuous variables is made via a conditioning on neighborhoods) can also become an attractive variant for binary data as demonstrated by Besag & Tantrum (2003).

Modeling and inference in metapopulation models has received considerable attention over the last 10 or 15 years. Much of the work has been on devising models of extinction and colonization, assuming the occupancy state was perfectly observable. The Markovian state model (without explicit spatial dynamics) was developed by Clark & Rosenzweig (1994). Hanski (1994), Day & Possingham (1995) and others have addressed spatial dynamics. Formalization of inference procedures has been addressed by Moilanen (1999) while O'Hara *et al.* (2002) and Ter Braak & Etienne (2003) provided a Bayesian treatment of the inference problem for occupancy models with temporal dynamics. These papers focus on the state process model and inferences under that model assume that the state-variable can be observed perfectly.

In metapopulation studies, when there is population turnover, it is necessary to resort to modeling approaches assuming many habitat patches and local populations. Among these approaches, Hanski (1994) and Hanski & Simberloff (1997) distinguished spatially implicit, spatially explicit and spatially realistic approaches. Spatially implicit approaches correspond to models in which all habitat patches and local populations are assumed to be equally connected to each other (e.g. Levins, 1970 ; Pulliam, 1988 ; Harrison & Quinn, 1989 ; Hanski & Gyllenberg, 1993). However, at some point, it is needed to incorporate specific information on the spatial locations of populations. In the scope of spatially explicit approaches, there are several modeling frameworks, such as cellular automata models (Caswell & Etter, 1993), interacting particle systems (Durrett, 1989), and coupled map lattice models (Hassell *et al.*, 1991). Here patches as arranged as cells on a regular grid (lattice) and populations are assumed to interact only with populations in the nearby "cells". The spatially

realistic approach includes in models the specific geometry of particular patch networks (e.g. number of patches in the network, location of these patches) (Hanski & Simberloff, 1997).

Limits of classical models – Spatial models for binary processes (sometimes referred to as "zero/one processes" or "presences/absence processes" or "occupancy processes") have a long history in the statistics literature (e.g., for a review see Cressie, 1993). Although such approaches have recently been extended to the spatio-temporal context (e.g., Zhu *et al.*, 2005), they have not focused on dynamic processes per se, and have not been used in the context of data with observational uncertainty. Specifically, the classical approaches to the development of such models are deficient for three reasons. First, they do not explicitly accommodate sampling uncertainty in the form of false absences. That is, observed zeros (putative absence) may arise because individuals are either absent from the site (a structural zero) or because the species was present but went undetected in the sampling (a sampling zero). Hierarchical modeling accounting for uncertainty in detection permits consideration of this problem, but even in that case, there are problems in accounting for sampling errors (Sauer *et al.*, 1994) or in conditioning model on the presence of the species of interest in the area where the sampling protocol is conducted. Secondly, there is a critical need for spatio-temporal models of occupancy, especially in the framework of hierarchical modeling. Even if recent efforts have been made to achieve a conceptual unification of models that are dynamic and models with a spatial influence on the dynamics, (Zhu *et al.*, 2005 ; Hooten & Wikle, 2008), there has not been a great deal of work on statistical modeling of spatio-temporal occupancy systems (Royle & Dorazio, 2008). Finally, existing models do not explicitly incorporate (i.e., in the model parameterization) an explicit linkage between data and population demographic process (e.g., recruitment, survival, migration, emigration). Most of the models are “fitting models” (i.e. phenomenological) and do not explicitly consider underlying mechanisms (Bennett *et al.*, 2001). However, it is obvious that when it is possible to use ecological mechanisms to describe and predict site occupancy, this is an advantage.

Classical issues in site occupancy models

Detectability – Virtually all ecological sampling processes for species occurrence data admit two possible events that can give rise to observed species absences: true absence, and presence but nondetection. However, virtually all ecological studies of species occurrence treat nondetections as true absences, leading to misleading inferences. Methods for modeling

and inference from such data in the presence of sampling and process uncertainty have only recently become a central focus of active research and development (MacKenzie *et al.*, 2005). Nevertheless, since these methods are relatively new, they have seen little use in macroecological investigations (Royle & Dorazio, 2008). More importantly, the current status of these methods renders them inefficient for modeling spatially and temporally dynamic processes.

Inferences about occupancy may be misleading when detection probability is not incorporated into the methods of data analysis. Not only will naïve approaches underestimate occupancy, but indices intended to reflect relative occupancy may also be biased (MacKenzie, 2006) and the effect of any variable of interest misidentified, particularly if detection probability covaries with the factors or variables thought to affect occupancy (Gu & Swihart, 2004 ; MacKenzie, 2006). Inferences about the dynamic processes that drive changes in occupancy may also be inaccurate (Moilanen, 2002 ; MacKenzie *et al.*, 2003). Therefore, robust inference about occupancy and related dynamics can only be made by explicitly accounting for detection probability. (MacKenzie *et al.*, 2005).

One important extension of models of occurrence or occupancy in the presence of imperfect detection is to the situation in which a site's occupancy status may change through time, i.e., to the situation in which the metapopulation system is "open" to local extinction and colonization events. MacKenzie *et al.* (2003) provided a general characterization of open models, and described a likelihood-based framework for inference about model parameters, while Royle & Kéry (2007) provided the corresponding hierarchical formulation. An site occupancy hierarchical model is described in Dupuis & Joachim (2006). The sampling design required to fit such models is commonly referred to as the robust design. (Pollock 1982; Kendall et al. 1995), in which replicate samples are made at each site subject to closure, and sampling is repeated over time. Under these open models, the metapopulation system is assumed to be closed within, but not across primary periods. Such models are referred as dynamic occupancy models.

If we consider data obtained from repeated presence/absence (more precisely detection/nondetection) surveys of i=1, 2, …, R spatial units (sites), and if we suppose that each site is surveyed k=1, 2, …, K times within each of t=1, 2, …, T *primary* periods and that each site is closed with respect to its occupancy status within but not across primary periods. A typical case would be surveys repeated several times both within the breeding season of a species and over several years. This situation is that for which the "robust design" (Pollock,

1982 ; Kendall *et al.*, 1995 ; Williams *et al.*, 2002) has been developed in conventional capture-recapture applications. Let $z_{i,t}$ denote the *true* occupancy status of unit *i* during primary period *t*, having possible states "occupied" ($z_{i,t}=1$) or "not occupied" ($z_{i,t}=0$) and $y_{i,t}$ the *observed* occupancy status. One parameter of interest is the probability of site occupancy (or the probability of occurrence) for period *t*, $\psi_t=Pr(z_{i,t}=1)$. Let φ_t be the probability that an occupied site "survives" (i.e., remains occupied) from period *t* to *t+1*, i.e., $\varphi_t=Pr(z_{i,t+1}=1|z_{i,t}=1)$. In metapopulation systems, *local colonization* is the analog of the recruitment process. Let γ_t be the local colonization probability from period *t* to *t+1*, i.e. $t=Pr(z_{i,t+1}=1|z_{i,t}=0)$. MacKenzie *et al.* (2003) consider a classical likelihood formulation of this problem in which the model does not contain the binary state-variables ($z_{i,t}$), just the survival and colonization probability parameters that govern the state process, φ_t and γ_t, respectively. That is, the latent indicators of occupancy are removed from the likelihood by integration. However, there are situations in which one is interested in the occupancy state and it is therefore important to consider a formulation that accommodates the prediction of these random effects. Although not considered as such in the ecological literature, this model can be naturally formulated in a state-space representation, in which the model is expressed by two component processes: a submodel for the observations conditional on the unobserved state process, i.e., $y_{i,t}|z_{i,t}$, and, secondly, a submodel for the un- or partially observed state process involving the detection probability $p_{i,t}$.

Therefore, we have the state process model :

$$Z_{i,t} \mid z_{i,t-1} \sim Bern(\psi_{i,t}) \text{ with } \psi_{i,t} = \Pr(Z_{i,t} = 1 \mid Z_{i,t-1} = z_{i,t-1}) = z_{i,t-1}\phi_{t-1} + (1 - z_{i,t-1})\gamma_{t-1}$$

and an observational process: $y_{i,t} \sim Bin(K_{i,t}, p.z_{i,t})$.

Despite the fact that this model and its extensions (e.g. MacKenzie *et al.*, 2005) are a huge improvement for occupancy modeling, some problems persist. The two main issues that we will deal with in this thesis arise from two distinct elements: an ambiguous situation in the implicit conditioning underlying this model, and a practical issue related to errors during the sampling protocol. First, in the case of a closed environment (in term of geographical delimitation), whether the probability that a species is detected in the sampled spatial units is conditioned on the presence of this species in the whole area of studied (i.e. sampled and unsampled units) is unclear. This issue is evocated in papers by Dupuis *et al.*, and Bled *et al.* (b.) focusing on the relationship between climatic factors and bird community state variables. Specifically, conditioning (or not) on the presence of species of interest at the global scale (i.e. the whole area of study containing both sampled and unsampled quadrats) is important

(Dupuis & Joachim, 2006). Typically, not accounting for this issue when species are detected only in few quadrats (which is typically the case of spatially rare species) might have an important impact on occupancy estimation. In this specific context, we have worked using an informative Bayesian approach, that permitted to study how the consideration of some a priori information on a parameter (e.g. detection probability) -which is not the parameter of interest (e.g. occupancy probability)- might improve the precision of the estimation of the quantity of interest. Secondly, many errors may arise from the data gathering, due to observers' inexperience or possible misidentifications and confusion with other species for example. This specific issue is addressed in the article about the propagation of an invasive species in the U.S.A.

One consequence of detection probability lower than 1 is the presence of missing data. The treatment of missing data has been an issue in statistics for some time, and it has come even more to the fore in recent years. As detailed for instance in McLachlan & Peel (2000) for mixture models or in Robert & Casella (1999) in a more general perspective, Markov Chain Monte Carlo (MCMC) methods have been deeply instrumental in the Bayesian exploration of increasingly complex missing data problems, as further shown by the explosion in the number of papers devoted to specific missing data models since the early 1990's (Celeux *et al.,* 2005). The current interest in missing data stems mostly from the problems caused in surveys and census data, but the topic is actually much broader than that. (Howell, 2008). Missing data create difficulties in scientific research because most data analysis procedures were not designed for them. Difficulties include the computation of likelihood and Bayesian estimations of quantity of interest. Missingness is usually a nuisance, not the main focus of inquiry, but handling it in a principled manner raises conceptual difficulties and computational challenges. Lacking resources or even a theoretical framework, researchers, methodologists, and software developers resort to editing the data to lend an appearance of completeness. Unfortunately, ad hoc edits may do more harm than good, producing answers that are biased, inefficient (lacking in power), and unreliable. (Scafer & Graham, 2002). The mechanisms (distribution patterns) of missing data are traditionally divided into three classes: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) (Rubin, 1976). If missing data are not MCAR, then there are potential problems in analyzing data as though they were, but the precise outcome depends on the way in which they are missing, specifically whether data are MAR or MNAR (Nakagawa & Freckleton, 2008). Therefore, when dealing with missing data, one should determine in the first place what type of missing he is facing, in order to choose the

appropriate way of dealing with such problem (see for example Little & Rubin, 1987 ; Allison, 2001 ; Schafer & Graham, 2002 or Howell, 2008 for reviews of such methods). Conventional methods of handling missing data include listwise deletion, pairwise deletion, dummy-variable adjustment, imputation, or maximum likelihood (Allison, 2001). A key component of dealing with missing data in the context of maximum likelihood (and therefore in Bayesian framework), whatever the class of distribution patterns is, is the use of data augmentation. Data augmentation is a widely used algorithm particularly in Markov chain Monte Carlo (MCMC) methods. The idea refers to methods for constructing sampling algorithms via the introduction of unobserved data or latent variables (Tanner & Wong, 1987 ; van Dyk & Meng, 2001) and arises naturally in missing value problems. Here, one takes advantage of the missing data structure, and this structure is used to compute a complete data likelihood where the missing data are simulated conditionally on the observed data. Such an approach is developed here in chapter IV in order to deal with missing data due to species with a detection probability inferior to 1. Examples of applications of data augmentation in a Bayesian framework include Dupuis (1995) (with the use of CMR data) and Dupuis & Schwarz (2007) where computations' difficulties for Bayesian estimation, hierarchical modeling, Gibbs sampling and the use of an informative a priori are also considered. Dupuis & Joachim (2003) have displayed the structure of missing data in data handled in the framework of site occupancy modeling, and the implementation of data augmentation in this context.

Spatio-temporal dimension – The distribution and abundance of organisms involves both physical and biotic processes that vary spatially and temporally, and are typically non-linear and dynamic. Methods for modeling and inference from occupancy data in the presence of sampling and process uncertainty have only recently become a central focus of active research and development. However, the current status of these methods renders them ineffective for modeling realistic spatio-temporal dynamic processes. Moreover, when the three quantities of potential interest in population ecology and management (abundance, occupancy and species richness) are investigated with respect to their distribution over space at one point in time, inferences about dynamics processes that produce spatial patterns are always very weak, as many alternative hypotheses can be invoked to explain most ecological patterns (MacKenzie *et al.*, 2005).

In ecological systems, concern in occupancy stems from the fundamental interest in the nature and strength of relations among and within local populations. Understanding spatial dynamics such as those driving the spread of invasive species (or diseases) is very important , as are intrinsic population demographic processes having to do with competition, recruitment, dispersal, source/sink dynamics, dynamics of species distribution and range, and other macroecological concepts. Despite this interest, attention to spatio-temporal dynamical models for occupancy has lagged behind, although there is a growing literature on spatial models for Gaussian processes in statistics (e.g., see Banerjee *et al.*, 2004 for a recent overview) and a well-developed literature on spatial models for binary processes (e.g., see Cressie, 1993 for an overview; Hoeting *et al.*, 2000). There have been recent efforts in ecology toward developing and using temporally dynamic models of occupancy (Barbraud *et al.*, 2003 ; MacKenzie *et al.*, 2003, Eraud *et al.*, 2007 ; Hooten *et al.*, 2007 ; Royle & Kery, 2007 ; Hooten & Wikle, 2008) that accommodate both explicit concepts of population dynamics as well as observation uncertainty due to sampling. Barbraud *et al.* (2003) dealt with spatio-temporal dynamics by modeling the colonization parameter for one region as a function of the local extinction parameter in a neighboring region. Other than these attempts, I know of no serious work on developing models with explicit spatial or spatio-temporal dynamics.

An interesting class of models is models that are motivated by viewing the movement (i.e., dispersal) of a phenomenon (e.g. invasion of an area, range shift) from the perspective of the phenomenon itself, rather than the system as a whole (i.e. directly try to model movement between each specific locations rather than a global population migration). For example, an exotic invasive species or a pathogenic species, in a new environment will often move from areas of lower quality to areas of higher quality (quality defined in terms of many possible factors, from environmental suitability to overpopulation through availability of hosts). Such models have connections to cellular automata models (e.g. Wolfram, 1984). Automata are most often defined in a deterministic dynamical system framework where the state of the "neighborhood" of an entity (in the case of an areal unit of space, called a "cell") determines the future state of the entity. Another type of automata can be formulated probabilistically, where movement of an entity to its neighborhood is defined by parametric probability distributions and the behavior of the system as a whole (i.e. the "automaton") -given the probability rule- is not unique, but can be expressed in terms of likelihood (Lee *et al.*, 1990). A propagating automatous system defined in either manner is capable of exhibiting spatially irregular wave-like behavior commonly found in natural phenomena (Hogeweg, 1988). Using

more traditional terminology found in the spatial statistics literature, "spatially irregular" refers to spatial structure that is either anisotropic (i.e., varying directionally) or non-stationary (i.e., varying locationally) or both (e.g. Cressie, 1993 ; Smith *et al.*, 2002). This fundamental difference in the construction of models is known as the "top-down" versus "bottom-up" approach (e.g. Grimm *et al.*, 2005). Traditionally, bottom-up approaches to studying ecosystem functioning and ecological processes have been used in simulation settings whereas top-down approaches have been taken using statistical methodology. Both have contributed much to ecology in theory and application, though they rarely intermingle.

In our paper on the propagation of an invasive species in USA, we develop a general hierarchical spatio-temporal dynamic occupancy framework accommodating the probabilistic automata case and generalizing it to the realistic situation in which there is uncertainty in the observations.

Another problem arises from the way occupancy is modeled at the global scale in the case of a closed environment. As a matter of fact in MacKenzie *et al.* (2005) the probability of presence at the global scale (the whole area studied) is completely determined by the one at the local scale (sampled quadrats in the whole area), and vice versa. Dupuis & Joachim (2006) have proposed a model which has two independent parameters, one by scale. The MacKenzie approach is suitable when the link between local and global occurrence probability seems reasonable. The model proposed by Dupuis & Joachim should be used in the absence of information about a possible link between these two quantities. We considered this distinction in the models proposed in papers by Dupuis *et al.* and its application to the study of the relationship between climatic changes and the occupancy rates of several species in a bird community in Bled *et al.* (b).

Fitting models and mechanistic models – Despite the intense recent interest in macro- and landscape- ecological questions (see examples below), investigations have brought to the forefront a methodological issue that can affect inference. This issue is the focus on pattern rather than process (Mackenzie *et al.*, 2005). The basic idea underlying such work is that current spatial patterns (e.g., of species distributions) carry information about the processes that generated them. While this idea is true to some extent, we note that it is possible to develop a large number of plausible hypotheses to explain any observed pattern (Phaedrus' Law *in* Pirsig, 1974 evocated in Nichols, 1991). Certainly some macroecological work has focused directly on processes (e.g., Boulinier *et al.*, 1998 ; Doherty *et al.*, 2003), but the vast majority of macroecological research has involved pattern.

Two general approaches to observational studies are used, and they are distinguished by the existence of *a priori* hypotheses. The observational studies that tend to be most useful to science are those in which conditional *a priori* hypotheses are specified and used to guide monitoring program design (Nichols, 2001 ; Williams *et al.*, 2002). Here different hypotheses predict different relationships between suspected causal factors and system state variables and specific predictions then emerge as changes in the causal factors occur naturally and are observed (MacKenzie *et al.*, 2005). The other approach involves the development of *a posteriori* hypotheses to explain observed system dynamics. Here data are fitted to various environmental and management variables in correlation and regression analyses in order to investigate possible relationships between population size or dynamics and these variables (MacKenzie *et al.*, 2005). The problem with this approach is that it is unlikely to yield "reliable knowledge" (Romesburg, 1981) because there will be typically multiple *a posteriori* hypotheses that provide reasonable explanations for any observed time series (Nichols, 1991). This problem becomes even more prominent if we consider the case of conducting association analyses of two time series and such analyses frequently lead to inappropriate inferences (Yule, 1926 ; Barker & Sauer, 1992).

From a scientific perspective, it is often of interest to make inference about the propagating nature of the process not only in the dynamics, but also in the covariance between state dynamics parameters and environmental heterogeneity or individuals behavioral choices. In the situation where a phenomenon is propagating in space over time, it is reasonable to think that certain areas of the spatial domain are more suitable than others. The ecological literature refers to this notion as "habitat suitability" or "habitat preference". From a dynamic modeling perspective it seems natural to think that a phenomenon will most likely progress from areas of undesirable habitat to areas of desirable habitat. Note that the term "habitat" is used very generally here, and refers to the successful propagation of the phenomenon under study. For example, if the phenomenon were a wildfire, it might "prefer" an exposed topography with steep slopes; whereas if it were a songbird, it might "prefer" forest edges.

In the paper focusing on nesting sites in the kittiwake by Bled *et al.* (c), we have proposed a model that explicitly encompasses hypotheses developed in the framework of habitat selection to estimate nesting site occupancy dynamics within a cliff.

Particular biological topics addressed in this thesis

The development of hierarchical spatio-temporal dynamic occupancy models is critical and timely as it provides a formal framework for considering inference and prediction of many ecological processes that directly affect the public welfare. These include species range modifications due to land use and climate changes, the spread of invasive species, and the spread of disease through animal vectors.

We have developed spatially and temporally dynamic models of occupancy that include explicit dynamic mechanisms. We have provided a general hierarchical framework that accommodates uncertainty in observations, process, and parameters. The primary goal of this research is to develop scientifically-meaningful, but statistically rigorous models for spatio-temporal dynamic occupancy processes in the presence of observation, model, and parameter uncertainty that can be used at several different geographical scales.

We wanted to propose a structure of models adaptable to several topics, correcting issues that appeared in previous general occupancy models. First of all, we were interested in the modeling of an invasive spread on North-American continent, the European Collared Dove, *Streptopelia decaocto*. We applied a model similar to the one described in this paper to the case of habitat selection within a cliff by a long-lived seabird, the kittiwake *Rissa tridactyla*, including hypotheses developed in the framework of habitat selection. Then, we address the issue of conditioning on the presence of the species in the area of interest and its implication for estimation of occupancy rate within sampled quadrats and over the whole area. Finally, using this model, we estimated the impact of climatic changes on the occupancy rates of several bird species in a French forest, considering spatial heterogeneity.

European Collared Dove: invasion process in the USA – The diversity and integrity of natural systems, as dynamic processes subject to various forces, can experience rapid changes and are particularly vulnerable to the presence of non-native organisms (Drake *et al.*, 1989). Exotic invasive species are often capable of out-competing native organisms for resources, utilizing broader food sources, interbreeding with and transmitting diseases to resident populations. Such invasions can be catastrophic in sensitive natural settings and will often result in extinctions, compromised genetic integrity, and direct impacts on human health and economy.

Considerable efforts have gone into the development of theoretical models of invasion spread (Hastings, 1996; Shigesada & Kawasaki, 1997 ; Hastings *et al.*, 2005). These models

have considered spread as an emergent process that arises from coupling population growth with movement; any process that influences growth or movement can be expected to affect rates of spread. These models have considerably enhanced our understanding of the spread process but empirical analyses are also needed. Analysis of patterns of spread in historical records for individual species provides critical insight into the spread process (Liebhold & Tobin, 2008). Of particular importance is the question of how the habitat effects spread and whether spread has been faster in areas where habitat characteristics promote either population growth or movement. (Morin *et al.*, 2009).

Numerous examples of invasions are well documented in the literature and occur in all Biotic Kingdoms; some of the most famous on the North American continent include: Kudzu, European Gypsy Moth, Japanese Beetle, Zebra Mussel, Africanized Honey Bee, and House Finch. Many other, less familiar, exotic species are just as detrimental to native systems and are rapidly propagating across the continent.

The Eurasian collared dove (*Streptopelia decaocto*) is one of the most successful invasive species, at least among terrestrial vertebrates (Romagosa & Labisky, 2000). This species has been introduced in North America during the early 1980's through Florida and is now detected as far away as the West Coast. (Dunn & Alderfer, 2006 ; Hooten & Wikle, 2008). A large dataset is available through data provided by the Breeding Bird Survey (B.B.S.). Moreover, we estimate spread direction, incorporate density dependant dynamics, detectability, and observer errors these data are subject to (Sauer *et al.*, 1994). In this work, we also consider the possibility of two distinct colonization processes through first and recolonization events.

The models we have developed is especially suited for modeling such spatio-temporal processes using binary data and spatially indexed covariates because of the direct manner in which we account for biological features of the invasion, such as persistence and long distance dispersal. Such a model could be used to identify important factors in the spread of epidemics (e.g., river corridors, shorelines, and population centers) and ultimately utilized in policy making and management decisions. Specific motivating examples are discussed below.

Nesting site selection by the kittiwake – Using data gathered in the framework of a survey led by Jean-Yves Monnat since 1979 (continuing at present), in six colonies of black-legged kittiwakes (*Rissa tridactyla*) located in Cap Sizun (France), we modeled nesting site use probability within a seabird subcolony (the kittiwake *Rissa tridactyla*). We integrated

hypotheses about habitat selection that predict how individuals select breeding habitat based on expected fitness in different habitat, and on local conditions of density and breeding success. We estimated demographic processes of site and colonization (once again through two subprocesses: first colonization and recolonization). Our model incorporated local and neighboring breeding success and conspecific density in the neighborhood: we addressed hypotheses about the influence of these factors on occupancy parameters.

Occupancy rates variations under climatic changes – As consequences of global warming, species are submitted to an augmentation of mean temperature and to the increased frequency of extreme temperatures. Species that are affected by these climatic variations might undergo modification in range and/or abundance. Using the improvement of estimates of occupancy rates provided by the work introduced in Dupuis *et al.*, we estimated modifications of occupancy rates in comparison with climatic changes of bird community. We have especially focused on how these variations are related to species preferential environmental conditions (i.e. inner forest vs edge), to differences between sedentary and migrating species, biogeographical origins of the different species. We have used data gathered by Jean Joachim in a South French forest after three particularly cold winters between 1985 and 1987.

The range or extent of occurrence of a species can be viewed as the primary element of the distributional component of ecology (Brown *et al.*, 1996) and has been termed .the basic unit of biogeography. (MacArthur, 1972). Conservation efforts require the ability to understand range and occupancy dynamics as a function of changes in dynamic features such as land use and climate. Such understanding will permit prediction of occupancy changes that are likely to accompany future changes and hopefully will permit informed attempts to mediate changes in occupancy that are viewed as undesirable. For example, spatial epidemiology is yet another area of investigation that can benefit from the methods introduced in this work, as failure to detect a disease or a parasite in an area of interest does not necessarily mean absence. In particular, we believe that the methods we develop will be useful for modeling the spread of important diseases in migratory bird populations (e.g., avian influenza and west nile virus, Marra et al. 2004) as well as in more sedentary wildlife populations (e.g., chronic wasting disease, Farnsworth *et al.*, 2006).

References

Anthony, R.G., Forsman E.D., Franklin A.B., Anderson D.R., Burnham K.P., White G.C., Schwarz C.J., Nichols J.D., Hines J.E., Olson G.S., Ackers S.H., Andrews S.L., Biswell B.L., Diller L.V., Dugger K.M., Fehring K.E., Fleming T.L., Gerhardt R.P., Gremel S.A., Gutierrez R.J., Happe P.J., Herter D.R., Higley S. M., Horn R.B., Irwin L., Loschl P.J., Reid J.A., & Govern S.G., 2006. Status and trends in demography of northern spotted owls, 1985–2003. *Wildlife Monographics*.

Banerjee S., Carlin B.P., Gelfand A.E., 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC.

Barbraud C., Nichols J. D., Hines J. E. & Hafner H., 2003. Estimating rates of extinction and colonization in colonial species and an extension to the metapopulation and community levels. *Oikos*. **101**: 113—126.

Barker R.J. & Sauer J.R., 1992. Modeling population change from time series data. In: *Wildlife 2001: Populations*. (N. Cappucino & P.W. Price, *eds.*) 182—194. Elsevier Applied Science, New York, USA.

Bell G., 2001. Neutral macroecology. *Science*. **293**: 2413—2418.

Bennett R.E., Nichols J.D., Lebreton J.D., Pradel R., Hines J.E. & Kitchen W.M., 2001. Methods for estimating dispersal probabilities and related parameters using marked animals. In: *Dispersal* (Clobert J., Danchin E., Dhondt A.A. & Nichols J.D., *eds*). 3 – 17. Oxford University Press.

Besag J. & Tantrum J., 2003. Likelihood analysis of binary data in space and time. In: *Highly structured stochastic systems* (P. Green, N. Hjort & S. Richardson, *eds.*). 289—295. New York: Oxford University Press.

Bled F. Royle J.A. & Cam E. a. Hierarchical modeling of an invasive spread: case of the European collared-dove *Streptopelia decaocto* in the USA. *Ecological Applications.In press.*

Bled F., Joachim J. & Dupuis J. b. Impact of climatic variations on bird species occupancy rate in a southern European Forest.

Bled F., Royle J.A. & Cam E. c. Assessing hypotheses about nesting site occupancy dynamics. *Ecology. In press.*

Boulinier T., Nichols J. D., Hines J. E., Sauer J. R., Flather C. H. & Pollock K. H., 1998. Higher temporal variability of forest breeding bird communities in fragmented landscapes. *Proceedings of the National Academy of Sciences*. **95**: 7497—7501.

Brown J. H., 1995. *Macroecology*. University Chicago Press, Chicago, IL, USA.

Brown J.H. & Maurer B.A., 1989. Macroecology: the division of food and space among species on continents. *Science*. **243**: 1145—1150.

Brown J.H., Stevens G.C. & Kaufman D.M., 1996. The geographic range: size, shape, boundaries, and internal structure. *Annual review of ecological systems*. **27**: 597—623.

Caswell H. & Etter R.J., 1993. Ecological interactions in patchy environments: From patch-occupancy models to cellular automata. In: *Patch Dynamics* (S.A. Levins, T.M. Powell and J.H. Steele, *eds.*) 93—109. Springer-Verlag, Berlin.

Celeux G., Marin J.-M. & Robert C.P., 2005. Iterated importance sampling in missing data problems. *Computational Statistics & Data analysis*. **50(12)**: 3386—3404.

Chessel D. & Thioulouse J., 2003. Analyse de données spatialisées. *Biostatistique*. Université de Lyon 1. (http://www.fr.embnet.org/R/pdf/bsa.pdf)

Clark C.W. & Rosenzweig M.L., 1994. Extinction and colonization processes: parameter estimates from sporadic surveys. *The American naturalist*. **159**: 351—362.

Connell J.H., 1963. Territorial behavior and dispersion in some marine invertebrates. *Research in Population Ecology*. **5**: 87—101.

Cressie N.A.C., 1993. *Statistics for Spatial Data*. Wiley.

Crow J.F. & Kimura M., 1970. *An introduction to Population Genetics Theory*. Harper & Row, New York, USA.

Day J.R. & Possingham H.P., 1995. A stochastic metapopulation model with variability in patch size and position. *Theoretical population biology*. **48**: 333—360.

Diamond J. M., 1975. Assembly of species communities. In: *Ecology and Evolution of Communities* (M. L. Cody and J. M. Diamond, *eds.*) 342—444. Harvard University Press, Cambridge, MA, USA.

Dice L.R., 1945. Measures of the amount of ecologic association between species. *Ecology*. **26**: 297—302.

Diggle P., 2003. *Statistical analysis of spatial point patterns* (2nd ed.) London: Arnold.

Doherty P.F. Jr., Boulinier T. & Nichols J.D., 2003. Extinction rates at the center and edge of species' ranges. *Annales Zoologici Fennici*. **40**: 145—153.

Drake J.A., DiCastri F., Groves R.H., Kruger F.G., Mooney H.A., Rejmanek M. & Williamson M.H. (*eds.*), 1989. *Biological Invasions: a global perspective*. John Wiley and Sons, New York, USA.

Dunn J.L. & Alderfer J., 2006. National Geographic Field Guide to the Birds of North America, Fifth Edition. *National Geographic Society, Washington*. 504p.

Dupuis J. & Joachim J., 2003. Bayesian estimation of species richness. *Publications du Laboratoire de statistique et probabilités.* **10:** 1—17.

Dupuis J. & Joachim J., 2006. Bayesian estimation of species richness from quadrat sampling data in the presence of priori information. *Biometrics*. **62**: 706 – 712.

Dupuis J. & Schwarz C. J., 2007. A Bayesian approach to the Multistate Jolly-Seber capture-recapture model . *Biometrics*. **63**: 1015-1022.

Dupuis J., 1995. Bayesian estimation of movement and survival probabilities from capture-recapture data. *Biometrika*. **82:** 761-772.

Dupuis J., Bled F. & Joachim J. Estimating occupancy rates of spatially rare or hard to detect species: a conditional approach. *Biometrics. In press.*

Durrett R., 1989. *Lecture notes on particle systems and percolation*. Wadsworth and Brooks/Cole, Pacific Grove, CA, USA. 326p.

Eraud C., Boutin J.-M., Roux D. & Faivre B., 2007. Spatial dynamics of an invasive bird species assessed using robust design occupancy analysis: the case of the Eurasian collared dove (*Streptopelia decaocto*) in France. *Journal of Biogeography*. **34(6)**: 1077—1086.

Fahrig L. & Merriam G., 1985. Habitat patch connectivity and population survival. *Ecology*. **66**: 1762—1766.

Farnsworth M.L., Hoeting J.A., Hobbs N.T. & Miller M.W., 2006. Linking chronic wasting disease to mule deer movement scales: a hierarchical Bayesian approach. *Ecological Applications*. **16**: 1026—1036.

Fisher R.A., 1937. The wave of advance of advantageous genes. *Annals of Eugenics*. **7**: 355—369.

Forbes S.A., 1907: On the local distribution of certain Illinois fishes. An essay in statistical ecology. *Bulletin of the Illinois State Laboratory of Natural History*. **7**: 273—303.

Gaston K.J. & Blackburn T.M., 2000. *Pattern and process in macroecology*. Blackwell, Malden, MA, USA.

Geary R.C., 1954. The contiguity ratio and statistical mapping. *The incorporated statistician*. **5**: 115—145.

Grimm V., Revilla E., Berger U., Jeltsch F., Mooij W., Railsback S., Thulke H.-H., Weiner J., Wiegand T. & DeAngelis D., 2005. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*. **310**: 987—991.

Gu W. & Swihart R.K., 2004. Absent of undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological conservation*. **116**: 195—203.

Hanski I. & Gyllenberg M., 1993. Two general metapopulation models and the core-satellite species hypothesis. *The American Naturalist*. **142**: 17—41.

Hanski I. & Simberloff D., 1997. The metapopulation approach, is history, conceptual domain, and application to conservation. In: *Metapopulation biology: Ecology, genetics, and Evolution* (I. Hanski I. and M.E. Gilpin, *esd.*) New York: Academic Press.

Hanski I., 1994. Patch-occupancy dynamics in fragmented landscapes. *Trends in Ecology and Evolution*. **9**: 131—135.

Hanski I.A. & Gilpin M.E., 1997. *Metapopulation Biology : Ecology, Genetics & Evolution.* Academic Press. San Diego, California, USA. 512p.

Harrison S. & Quinn J.F., 1989. Correlated environments and the persistence of metapopulations. *Oïkos*. **56**: 293—298.

Hassell M.P., Comins H.N. & May R.M., 1991. Spatial structure and chaos in insect population dynamics. *Nature*. **353**: 255—258.

Hastings A., 1996. Models of spatial spread: is the theory complete? *Ecology*. **77**: 1675—1679.

Hastings A., Cuddington K., Davies K.F., Dugaw C.J., Elmendorf S., Freestone A., Harrison S., Holland M., Lambrinos J., Malvadkar U., Melbourne B.A., Moore K., Caz Taylor C. & Thomson D., 2005. The spatial spread of invasions: new developments in theory and evidence. *Ecology Letters*. **8**: 91—101.

Hoeting J., Leecaster M. & Bowden D., 2000. An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics*. **5**: 102—114.

Hogeweg P., 1988. Cellular automata as a paradigm for ecological modeling. *Applied Mathematics and Computation*. **27**: 81—100.

Hooten M.B, Wikle C.K, Dorazio R.M, Royle J.A., 2007. Hierarchical Spatiotemporal Matrix Models for Characterizing Invasions. *Biometrics*. **63** (2) : 558 – 567.

Hooten M.B. & Wikle C.K., 2008. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environmental and Ecological Statistics*. **15**: 59 – 70.

Hubbell S.P., 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton Univ. Press, Princeton, NJ.

Kareiva P., 1990. Population dynamics in spatially complex environments: Theory and data. In: *Population regulation and dynamics* (M.P. Hassell and R.M. May, *eds.*) 53—68. Royal Society, London, UK.

Kendall W.L., 2001. Using models to facilitate complex decisions. In: *Modeling in Natural Resource Management* (T.M. Shenk & A.B. Franklin, *eds.*). 147 – 170. Island Press, Washington, D.C., USA.

Kendall W.L., Pollock K.H. & Brownie C., 1995. A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics*. **51**: 293—308.

Kot M., 1992. Discrete-time travelling waves: ecological examples. *Journal of Mathematical Biology*. **30**: 413—443.

Krebs C., 1978. *Ecology: The Experimental Analysis of Distribution and Abundance*. Harper and Row Publishers, Inc., New York, New York, USA.

Lawson A.B., 2006. Statistical Methods in Spatial Epidemiology. *Wiley Series in Probability and Statistics*. 424 p.

Le Bras H., 1971. Equilibre et croissance des populations soumises à des migrations. *Theoretical Population Biology*. **2**: 100—121.

Lebreton J.-D. & Gonzalez-Davilla G., 1993. An introduction to models of subdivided populations. *Journal of Biological Systems*. **1**: 389—423.

Lebreton J.-D., 1996. Demographic models for subdivided populations: the renewal equation approach. *Theoretical Population Biology*. **49**: 291—313.

Lebreton J.-D., Khaladi M. & Grosbois V., 2000. An explicit approach to evolutionary stable dispersal strategies: No cost of dispersal. *Mathematical Biosciences*. **165**: 163—176.

Lee Y., Qian S., Jones R., Barnes C., Flake G., O'Rourke M., Lee K., Chen H., Sun G., Zhang Y. & Chen D., 1990. Adaptive stochastic cellular automata: Theory. *Physica D.* **45**: 159-180.

Levins R., 1970. Extinction. 75-107 In: *Some mathematical problems in biology* (M. Gerstenhaber, *ed.*). Providence, RI: American Mathematical Society.

Liebhold A.M. & Tobin P.C., 2008. Population ecology of insect invasions and their management. *Annual Reviews of Entomology*. **53**: 387—408.

Ludwig J.A. & Reynolds J.F., 1988. *Statistical ecology: a primer on methods and computing*. John Wiley & Sons, Inc. New York, USA. 352p.

MacArthur R.H. & Wilson E.O., 1963. An equilibrium theory of insular biogeography. *Evolution*. 17: 373—387.

MacArthur R.H. & Wilson E.O., 1967. *The theory of island biogeography*. Princeton University Press, Princeton.

MacArthur R.H., 1972. *Geographical ecology*. Harper and Row, New York, USA.

MacKenzie D.I., 2006. Modeling the probability of resource use: the effect of, and dealing with, detecting a species imperfectly. *Journal of Wildlife Management*. **70(2)**: 367—374.

MacKenzie D.I., Nichols J.D., Hines J.E., Knutson M.G. & Franklin A.D., 2003. Estimating site occupancy colonization and local extinction when a species is detected imperfectly. *Ecology*. **84**: 2200—2207.

MacKenzie D.I., Nichols J.D., Royle J.A., Pollock K.H., Bailey L.L. & Hines J.E., 2006. *Occupancy Estimation and Modeling*. Academic Press. San Diego, California, USA. 324p.

Marra P. P., Griffing S., Caffrey C., Kilpatrick A.M., McLean R., Brand C., Saito E., Dupuis A.P., Kramer L. & Novak R., 2004. West Nile Virus and wildlife. *Bioscience*. **54**: 393—402.

Matérn B., 1986. Spatial variation. **36**. *Lecture notes in statistics*. New York: Springer.

May R.M., 1976. Simple mathematical models with very complicated dynamics. *Nature*. **261**: 459—467.

McIntosh R.P., 1991. Concept and terminology of homogeneity and heterogeneity in ecology. *Ecological Studies*. **86**: 24—46.

McLachlan G. & Peel D., 2000. Finite Mixture Models. J. Wiley, New York.

Moilanen A., 1999. Patch occupancy models of metapopulation dynamics: Efficient parameter estimation using implicit statistical inference. *Ecology*. **80**: 1031—1043.

Moilanen A., 2002. Implications of empirical data quality to metapopulation model parameter estimation and application. *Oïkos*. **96**: 516—530.

Morin R.S., Liebhold A.M. & Gottschalk K.W., 2009. Anisotropic spread of hemlock woolly adelgid in the eastern United States. *Biological invasions*. DOI 10.1007/s10530-008-9420-1.

Neubert M.G. & Caswell H., 2000. Demography and dispersal: calculation and sensitivity analysis of invasion speed for structured populations. *Ecology*. **81(6)**: 1613—1628.

Nichols J.D., 1991. Science, population ecology, and the management of the American black duck. *The Journal of Wildlife Management*. **55(4)**: 790—799.

Nichols J.D., 2001. Using models in the conduct of science and management of natural resources. In: *Modeling in Natural Resource Management: Development, Interpretation and Application* (T.M. Shenl & A.B. Franklin, *eds.*) 11—34. Island Press, Washington D.C., USA.

Nichols, J. D., 2004. Evolution of quantitative methods for the study and management of avian populations : on the importance of individual contributions. *Animal Biodiversity and Conservation*. **27(1)**: 3–19.

O'Hara R.B., Arjas E., Toivonen H. & Hanski I., 2002. Bayesian analysis of metapopulation data. *Ecology*. **83**: 2408—2415.

Ortega-Cejas V., Fort J. & Méndez V., 2004. The role of the delay time in the modeling of biological range expansions. *Ecology*. **85**: 258—264.

Patterson B.D. & Atmar W., 1986. Nested subsets and the structure of insular mammalian faunas and and archipelagos. *Biological Journal of the Linnean Society*. **28**: 65—82.

Patterson B.D., 1987. The principle of nested subsets and its implications for biological conservation. *Conservation Biology*. **1**: 323—334.

Pearce J.L. & Boyce M.S., 2006. Modeling distribution and abundance with presence-only data. *Journal of applied ecology*. **43**: 405—412.

Pirsig R.M., 1974. *Zen and the art of motorcycle maintenance*. Bantam Books, New York.

Pollock K. H., 1982. A capture-recapture design robust to unequal probability of capture. *Journal of Wildlife Management*. **46**: 757—760.

Pulliam H.R., 1988. Sources, sinks, and population regulation. *The American Naturalist*. **132**: 652—661.

Ripley B.D., 1981. *Spatial statistics*. New York: Wiley.

Robert C.& Casella G., 1999. Monte Carlo Statistical Methods. Springer-Verlag, New York.

Rogers A., 1966. The multiregional matrix growth operator and the stable interregional age structure. *Demography*. **3**: 537—544.

Rogers A., 1968. *Matrix analysis of international population growth and distribution*. University of California Press, Berkeley, CA, USA.

Rogers A., 1975. *Introduction to multiregional mathematical demography*. Wiley, New York, USA.

Rogers A., 1985. *Regional population projection models*. Sage publications, Beverly Hills, CA, USA.

Rogers A., 1995. *Multiregional demography: principles, methods, and extensions*. Wiley, New York.

Romagosa C.M. & Labisky R.F., 2000. Establishment and Dispersal of the Eurasian Collared-Dove in Florida. *Journal of Field Ornithology*. **71(1)**: 159—166.

Romesburg H.C., 1981. Wildlife science: Gaining reliable knowledge. *Journal of Wildlife Management*. **45**: 293—313.

Rosenzweig M., 1995. *Species diversity in space and time*. Cambridge University Press.

Royle J.A. & Dorazio R.M., 2008. *Hierarchical Modeling and inference in ecology (The analysis of data from populations, metapopulations, and communities)*. Academic Press.

Royle J.A. & Kery M., 2007. A Bayesian state-space formulation of dynamic occupancy models. *Ecology*. **88(7)**: 1813—1823.

Royle J.A. & Nichols J.D., 2003. Estimating abundance from repeated presence–absence data or point counts. *Ecology*. **84(3)**: 777—790.

Royle J.A., Nichols J.D. & Kéry M., 2005. Modeling occurrence and abundance of species when detection is imperfect. *Oïkos*. **110**: 353—359.

Sauer J. R., Peterjohn B.G. & Link W.A., 1994. Observer differences in the North American Breeding Bird Survey. *The Auk*. **111(1)**: 50—62.

Schoen R., 1988. *Modeling multigroup populations*. Plenum Press, New York, USA.

Shigesada N. & Kawasaki K., 1997. *Biological invasions: theory and practice*. Oxford University Press, UK.

Skellam J.G., 1951. Random dispersal in theoretical populations. *Biometrika*. **38**: 196–218.

Smith D., Lucey B., Waller L., Childs J. & Real L., 2002. Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proceedings of the National Academy*. **99**: 3668-3672.

Ter Braak C.J. & Etienne R.S., 2003. Improved Bayesian analysis of metapopulation data with an applicatoin to a tree frog metapopulation. *Ecology*. **84**: 231—241.

Van den Bosch F., Hengeveld R. & Metz J.A.J., 1992. Analysing the Velocity of Animal Range Expansion. *Journal of Biogeography*. **19**(2): 135—150.

Van den Bosh F., Metz J.A.J. & Diekmann O., 1990. The velocity of spatial population expansion. *Journal of Mathematical Biology*. **28**: 529—565.

Van Dyk D.A & Meng X.-L., 2001. The art of data augmentation. *Journal of Computational and Graphical Statistics*. **10(1)**: 1—50.

Vitousek P.M., Mooney H.A., Lubchenco J. & Melillo J.M., 1997. Human domination of Earth's ecosystems. *Science*. **277**: 494—499.

Wikle C.K., 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*. **84(6)**: 1382—1394.

Williams B.K., Nichols J.D. & Conroy M.J., 2002. *Analysis and Management of Animal Populations*. Academic Press. San Diego, California, USA. 817p.

Wolfram S., 1984. Cellular automata as models for complexity. *Nature*. **311**: 419—424.

Yule G.U., 1926. Why do sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time-series. *Journal of the Royal Statistics Society*. **89**: 1—69.

Zhu, H., Huang H.-C. & Wu C.-T., 2005. Modeling spatial-temporal binary data using Markov random fields. *Journal of Agricultural, Biological, and Environmental Statistics*. **10:** 212 – 225.

ABSTRACTS

1

Hierarchical modeling of an invasive spread: case of the Eurasian collared-dove *Streptopelia decaocto* in the USA

Modélisation hiérarchique d'une propagation invasive: cas de la tourterelle turque Streptopelia decaocto aux Etats-Unis

Bled F.[1]*, Royle J.A.[2] & Cam E.[1]

[1] Laboratoire Évolution et Diversité Biologique, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 9, France.

[2] U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland 20708, USA.

(Ecological Applications, in press)

***** Corresponding author (bled@cict.fr)

Abstract – Invasive species are regularly claimed as the second threat on biodiversity. To apply a relevant response to the potential consequences associated with invasions (e.g. emphasize management efforts to prevent new colonization or to eradicate the species in places where it has already settled), it is essential to understand invasion mechanisms and dynamics. Quantifying and understanding what influences rates of spatial spread is a key research area for invasion theory.

In this paper, we develop a model to account for occupancy dynamics of an invasive species. Our model extends existing ones to accommodate several elements of invasive processes; we chose the framework of hierarchical modeling to assess site occupancy status during an invasion.

First, we explicitly accounted for spatial structure and how distance among sites and relative position to one another affect the invasion spread. We especially accounted for the case of directional propagation and provide a way of estimating the direction of this possible spread. Secondly, we considered the influence of local density on site occupancy. Thirdly, we decided to split the colonization process into two subprocesses, first colonization and recolonization, which may be ground-breaking since these subprocesses may exhibit different relationships with environmental variations (such as density variation) or colonization history (e.g. first colonization might facilitate further colonization events). Finally, our model incorporates imperfection in detection, which might be a source of a substantial bias in population parameters estimation.

We focused on the case of the Eurasian collared dove (*Streptopelia decaocto*) using data from the B.B.S. The Eurasian collared dove is one of the most successful invasive species, at least among terrestrial vertebrates, and its invasion of the U.S.A. since its introduction in the early 1980's. Our model provided estimation of the spread direction consistent with empirical observations. Site persistence probability exhibits a quadratic response to density. We also succeeded at detecting differences in the relationship between first colonization and recolonization probabilities and density. We also provide a map of future possibly colonized sites as an example of possible practical application of our work through.

Keywords: Hierarchical modeling, invasive species, detectability, first and re-colonization, site occupancy

1

Résumé – Les espèces invasives sont souvent présentées comme étant la seconde menace pour la biodiversité. En vue d'appliquer une réponse pertinente aux dangers potentiels associés aux invasions (e.g. accentuer les efforts de gestion sur le fait d'éviter de nouvelles colonisations ou sur l'éradication de l'espèce dans les endroits où elle est déjà présente), il est essentiel de comprendre les mécanismes d'une invasion et sa dynamique. La quantification et la compréhension de ce qui influence les taux de diffusion spatiale constituent un domaine clé de recherche pour la théorie d'invasion.

Dans ce papier, nous présentons un modèle qui décrit la dynamique d'occupation d'une espèce invasive. Notre modèle prolonge les modèles existants pour intégrer plusieurs éléments des processus d'invasion en utilisant ce qui a été développé dans le domaine de la modélisation hiérarchique pour modéliser l'occupation des sites au cours d'une invasion.

Dans un premier temps, nous prenons explicitement en compte la structure spatiale et incluons dans notre modèle la manière dont la distance entre les sites et leur position relative les uns par rapport aux autres affecte le processus d'invasion. Nous considérons plus spécialement le cas d'une diffusion directionnelle et fournissons une manière d'estimer une direction préférentielle de cette propagation. Puis, nous considérons l'influence de la densité locale sur l'occupation des sites. Troisièmement, nous séparons le processus de colonisation en deux sous-processus de première et recolonisation, ce qui pourrait être une avancée révolutionnaire puisque ces sous-processus pourraient réagir différemment aux variations environnementales (comme des changements de densité) ou à l'histoire de colonisation (e.g. une première colonisation pourrait faciliter de nouvelles colonisations). Enfin, nous incorporons dans notre modèle les problèmes de détectabilité, qui peut être une source de biais important dans l'estimation des paramètres de la population.

Nous nous sommes focalisés sur le cas de la tourterelle turque (*Streptopelia decaocto*) qui est une des espèces invasives les plus efficaces, au moins en ce qui concerne les vertébrés terrestres. Nous nous sommes intéressés à son invasion du territoire américain depuis son introduction au début des années 1980, en utilisant les données du B.B.S. Notre modèle nous a donné une estimation de la direction de propagation cohérente aux observations empiriques. La probabilité de persistance des sites montre une réponse quadratique à la densité. Nous sommes également parvenus à détecter une différence de réponse des probabilités de première colonisation et de recolonisation aux changements de densité. Nous fournissons également un exemple d'application pratique de notre travail au travers d'une carte indiquant les futurs sites ayant un fort risque d'être colonisés.

Mots-clés: Modélisation hiérarchique, espèces invasives, détectabilité, première et re-colonisation, occupation des sites.

Assessing hypotheses about nesting site occupancy dynamics

Evaluation d'hypothèses sur la dynamique d'occupation des sites de nidification

Bled F.[1]*, Royle J.A.[2] & Cam E.[1]

[1] Laboratoire Évolution et Diversité Biologique, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 9, France.

[2] U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland 20708, USA.

(Ecology, in press)

***** Corresponding author (bled@cict.fr)

Abstract

Hypotheses about habitat selection developed in the evolutionary ecology framework assume that individuals, under some conditions, select breeding habitat based on expected fitness in different habitat. The relationship between habitat quality and fitness may be reflected by breeding success of individuals, which may in turn be used to assess habitat quality. Habitat quality may also be assessed via local density: if high-quality sites are preferentially used, high density may reflect high-quality habitat.

Here we assessed whether site occupancy dynamics vary with site surrogates for habitat quality. We modeled nest site use probability in a seabird subcolony (the kittiwake *Rissa tridactyla*), over a 20 year period. We estimated site persistence (an occupied site remains occupied from time t to t+1), and colonization through two subprocesses: first colonization (site creation at the time scale of the study) and recolonization (a site is colonized again after being deserted). Our model explicitly incorporated site-specific and neighboring breeding success and conspecific density in the neighborhood. Our results provided evidence that reproductively "successful" sites have a higher persistence probability than "unsuccessful" ones. Analyses of site fidelity in marked birds and of survival probability showed that high site persistence predominantly reflects site fidelity, not immediate colonization by new owners after emigration or death of previous owners. There is a negative quadratic relationship between local density and persistence probability. First colonization probability decreases with density, whereas recolonization probability is constant. This highlights the importance of distinguishing initial and recolonization to understand site occupancy. All dynamics varied positively with neighboring breeding success. We found evidence of a positive interaction between site-specific and neighboring breeding success.

We addressed local population dynamics using a site occupancy approach integrating hypotheses developed in behavioral ecology to account for individual decisions. This allows development of models of population and metapopulation dynamics that explicitly incorporate ecological and evolutionary processes.

Keywords: *Rissa tridactyla*, habitat selection, first and re-colonization probability, site occupancy, density, breeding success

1

Résumé –Les hypothèses sur la sélection de l'habitat supposent que les individus sélectionnent leur habitat de reproduction d'après la fitness attendue dans différents habitats sous certaines conditions. En plus de l'évaluation directe de la qualité de l'habitat au travers de critères physiques ou biotiques,, la relation entre la qualité de l'habitat et la fitness peut être partiellement reflétée par le résultat de la reproduction des individus, qui en retour peut être utilisé pour évaluer la qualité de l'habitat. Cette qualité peut également être évaluée au travers de la densité locale : si les sites de bonne qualité sont utilisés préférentiellement, une densité élevée pourrait réfléchir un habitat de bonne qualité.

Ici, nous avons évalué si la dynamique d'occupation des sites variait avec ces substituts de la qualité de l'habitat. Nous avons modélisé la probabilité d'utilisation des sites de nidification au sein d'une sous-colonie d'un oiseau marin longévif (la mouette tridactyle *Rissa tridactyla*) sur une période de 20 ans. Nous avons estimé la persistance (un site occupé reste occupé) et la colonisation des sites. La colonisation a été séparé en deux sous-processus: la première colonisation (création de site de nidification dans la base de données) et la recolonisation (réutilisation d'un site après un abandon). Notre modèle incorpore explicitement les succès de reproduction locaux et du voisinage, ainsi que la densité des congénères.

Nos résultats démontrent que les sites en succès ont une probabilité de persistance plus élevée que ceux en échec. Des analyses de la fidélité au site d'oiseaux marqués et des probabilité de survie ont montré qu'une persistance élevé reflète principalement la fidélité au site, et non une colonisation immédiate par de nouveaux propriétaires après une émigration ou mort des précédents propriétaires. Il y a une relation quadratique négative entre la densité locale et la probabilité de persistance. La probabilité de première colonisation diminue avec la densité, alors que ce n'est pas le cas de la probabilité de recolonisation. Ceci souligne l'importance de distinguer première et recolonisation si on souhaite comprendre l'occupation des sites. Tous les paramètres dynamiques varient positivement avec le succès de reproduction du voisinage. Nous avons montré l'existence d'une interaction entre succès de reproduction local et du voisinage.

Mots-clés: *Rissa tridactyla*, sélection de l'habitat, probabilité de première et de re-colonisation, occupation des sites, densité, succès reproducteur.

Estimating the occupancy rate of spatially rare or hard to detect species: a conditional approach

Estimation des taux d'occupation d'espèces spatialement rares ou difficilement détectables: une approche conditionnelle

Jérôme A. Dupuis[1,*], Florent Bled[2] and Jean Joachim[3]

[1] Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse, France

[2] Laboratoire Evolution et Diversité Biologique, Université Paul Sabatier, France

[3] INRA-CEFS, BP 27, 31326, Castanet-Tolosan, Toulouse, France

(Biometrics, in press)

***** Corresponding author (dupuis@cict.fr)

Summary – We consider the problem of estimating the occupancy rate of a target species in a region divided in spatial units (called quadrats); this quantity being defined as the proportion of quadrats occupied by this species. We mainly focus on spatially rare or hard to detect species which are typically detected in few quadrats, and for which estimating the occupancy rate (with an acceptable precision) is problematical. We develop a conditional approach for estimating the quantity of interest. The conditioning, which bears on the presence of the target species in the region of study, makes identifiable the occurrence and detectability parameters, regardless of the number visits made in the sampled quadrats. Compared with an unconditional approach, it proves to be complementary, in that this allows us to deal with biological questions which cannot be addressed by the former. Two bayesian analyses of the data are performed: one is non informative, and the other takes advantage of the fact that some prior information on detectability is available. It emerges that taking such a prior into account significantly improves the precision of the estimate when the target species has been detected in few quadrats and is known to be easily detectable.

Key words: Bayesian estimation; Identifiability; Missing data; Occupancy rate; Quadrat sampling; Spatially rare species.

Résumé – Nous nous intéressons à l'estimation des taux d'occupation d'une espèce d'intérêt dans une région divisée en sous-unités spatiales, appelées quadrats; cette quantité étant définie comme la proportion de quadrats occupés par cette espèce. Nous nous intéressons particulièrement aux espèces spatialement rares ou difficilement détectables, et qui sont typiquement détectées dans un faible nombre de quadrats, et pour lesquelles l'estimation du taux d'occupation (avec une précision acceptable) est problématique.

Nous avons développé une approche conditionnelle pour l'estimation de la quantité d'intérêt. Le conditionnement, qui porte sur la présence de l'espèce d'intérêt dans la région d'étude, rend identifiables les paramètres d'occurrence et de détectabilité, indépendamment du nombre de visites faites au sein des quadrat échantillonnés. Comparé à une approche non conditionnelle, notre approche est complémentaire, elle permet de traiter des questions biologiques qui ne peuvent être étudiées autrement.

Deux analyses bayésiennes des données sont effectuées : la première non informative, et l'autre tire avantage de la disponibilité d'une information a priori sur la détectabilité. Prendre en compte cette information a priori améliore significativement la précision des estimations lorsque l'espèce d'intérêt n'a été détectée que dans quelques quadrats et est facilement détectable.

Mots-clés: Estimation bayésienne ; Identifiabilité ; Données manquantes ; Taux d'occupation ; Échantillonnage par quadrat ; Espèces spatialement rares.

Impact of climatic variations on bird species occupancy rate in a southern European forest

Impact des variations climatiques sur les taux d'occupation d'espèces avicoles dans une forêt du Sud de la France

BLED Florent[1,*], JOACHIM Jean[2], DUPUIS Jérôme[3]

[1] Laboratoire Evolution et Diversité Biologique, Université Paul Sabatier, France
[2] INRA-CEFS, BP 27, 31326, Castanet-Tolosan, Toulouse, France
[3] Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse, France

(Biodiversity and Conservation, in press)

* Corresponding author (bled@cict.fr)

Abstract – Species that are affected by climatic variations can undergo modification in range and/or abundance. Knowing how individuals or species occupy their habitat is essential to understand how species use their environment, and detecting variations that might affect this use can be determinant in species management.

Hierarchical modeling is regularly used to assess for occupancy rate (i.e. proportion of patches occupied in a region), particularly when it is required to consider detectability-related issues. The present study is the first application of the conditional model presented in Dupuis *et al.* (2010), which is applied in the case of a heterogeneous area that might be divided into homogeneous sub-areas. Their approach is used to study the impact of three consecutive particularly cold winters on a selected set of bird species in a forest of southern France in the context of available prior information on birds detectability.

We examined a limited range of factors that might influence the response of some bird species to climatic. We considered the case of sedentary, partially migrating and migrating species. We also assessed if the biogeographical origins of the different species affect their occupancy rates. Globally, changes in occupancy rates between 1985 and 1987 indicates for the first time a continentalization of the regional forest fauna, reflected by the expansion of Palearctic and Turkestano-european faunistic type species, with depletion or extinction of European, Turkestano-mediterranean and Mediterranean sedentary species. We have also shown the importance of prior information.

Keywords: Climatic changes, detectability, hierarchical modeling, occupancy rate, prior information.

Résumé – Les espèces touchées par des variations climatiques peuvent subir des modifications de leur répartition géographique et/ou de leur abondance. Savoir comment les espèces occupent leur habitat est essentiel pour comprendre comment celles-ci utilisent leur environnement., et détecter les variations qui peuvent affecter cette utilisation peut être déterminant dans la gestion des espèces.

La modélisation hiérarchique est souvent utilisée pour estimer le taux d'occupation (i.e. la proportion de patchs occupés dans une région), particulièrement lorsqu'on doit considérer des problèmes liés à la détectabilité. Notre étude est la première application du modèle conditionnel présenté par Dupuis *et al.* (2010) qui est utilisé dans le cas d'une zone hétérogène qui peut être divisée en sous-ensembles homogènes. Ce modèle est utilisé pour étudier l'impact de 3 hivers consécutifs particulièrement rigoureux sur un ensemble d'espèces d'oiseaux dans une forêt du Sud de la France, dans le cas où une information a priori sur la détectabilité des espèces est disponible.

Nous nous focalisons sur un ensemble de facteurs qui pourraient influencer la réponse de certaines espèces au climat. Nous examinons le cas des espèces sédentaires, migratrices partielles et migratrices. Nous regardons également si les origines biogéographiques des différentes espèces peuvent affecter les variations des taux d'occupation. Globalement, l'évolution des taux d'occupations entre 1985 et 1987 indique pour la première fois une continentalisation de la faune forestière régionale, reflété par l'expansion des espèces de type faunistique paléarctique et turkestano-européen, avec une déplétion ou une extinction des espèces sédentaires de type européen, turkestano-méditrranéen et méditerranéen. Nous avons également pu mettre en évidence l'importance de l'information a priori.

Mots-clés: Changements climatiques, détectabilité, modélisation hiérarchique, taux d'occupation, information a priori.

CHAPTER II

1

Hierarchical modeling of an invasive spread: case of the Eurasian collared-dove *Streptopelia decaocto* in the USA

Bled F.[1]*, Royle J.A.[2] & Cam E.[1]

[1] Laboratoire Évolution et Diversité Biologique, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 9, France.

[2] U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland 20708, USA.

(Ecological Applications, in press)

***** Corresponding author (bled@cict.fr)

Abstract

Invasive species are regularly claimed as the second threat on biodiversity. To apply a relevant response to the potential consequences associated with invasions (e.g. emphasize management efforts to prevent new colonization or to eradicate the species in places where it has already settled), it is essential to understand invasion mechanisms and dynamics. Quantifying and understanding what influences rates of spatial spread is a key research area for invasion theory.

In this paper, we develop a model to account for occupancy dynamics of an invasive species. Our model extends existing ones to accommodate several elements of invasive processes; we chose the framework of hierarchical modeling to assess site occupancy status during an invasion.

First, we explicitly accounted for spatial structure and how distance among sites and relative position to one another affect the invasion spread. We especially accounted for the case of directional propagation and provide a way of estimating the direction of this possible spread. Secondly, we considered the influence of local density on site occupancy. Thirdly, we decided to split the colonization process into two subprocesses, first colonization and recolonization, which may be ground-breaking since these subprocesses may exhibit different relationships with environmental variations (such as density variation) or colonization history (e.g. first colonization might facilitate further colonization events). Finally, our model incorporates imperfection in detection, which might be a source of a substantial bias in population parameters estimation.

We focused on the case of the Eurasian collared dove (*Streptopelia decaocto*) using data from the B.B.S. The Eurasian collared dove is one of the most successful invasive species, at least among terrestrial vertebrates, and its invasion of the U.S.A. since its introduction in the early 1980's. Our model provided estimation of the spread direction consistent with empirical observations. Site persistence probability exhibits a quadratic response to density. We also succeeded at detecting differences in the relationship between first colonization and recolonization probabilities and density. We also provide a map of future possibly colonized sites as an example of possible practical application of our work through.

Keywords: Hierarchical modeling, invasive species, detectability, first and re-colonization, site occupancy

INTRODUCTION

Regularly claimed as the second threat on biodiversity and threatened species after habitat destruction (Glowka *et al.*, 1994 in Williamson, 1999 ; Wilcove *et al.*, 1998), invasive species are defined as species not naturally present in a geographic area that have been introduced by man, and that succeeded in establishing and colonizing this area. Even if only a small fraction of transported species become established and of these generally only about 1% become pests (Williamson, 1996), events of invasions are not uncommon. Invasive species are present in a wide range of taxa, from microorganisms to vertebrates, plants and invertebrates[*]. Over the last 500 years, invasive species have been estimated to have come to dominate 3% of the Earth's ice-free surface (Mack, 1985 *in* Mooney & Cleland, 2001). Moreover, just in Europe, 10670 species (fauna and flora) have been considered as "invasive exotic species" in the framework of the UE Daisie program.

While invasive species are generally presented as a strong threat to indigenous species, lots of studies on this subject are just correlative. They cannot conclusively determine if invasive species are responsible for the loss of biodiversity or just a result of what has caused this loss (e.g. habitat alteration) (Gurevitch & Padilla, 2004). Invasive exotic species are causing changes in many ecological systems worldwide, and there are altering many communities and ecosystems (Gurevitch & Padilla, 2004). However, since these species are not the only element affecting biodiversity and usually co-occur with other threats, it is essential to understand invasion mechanisms and dynamics in a wider context of global change to develop a relevant response to the potential consequences associated with invasions, e.g. emphasize management efforts to prevent new colonization (i.e. try to contain the invasion) or to eradicate the species in places where it has already settled. Quantifying and understanding what influences rates of spatial spread constitute a key research area for

[*] For an invasive species database, see the ISSG website - www.issg.org/database

invasion theory (e.g. Skellam, 1951 ; Okubo, 1980 ; Andow *et al.*, 1990; Kot *et al.*, 1996 ; Neubert & Caswell, 2000 ; Wikle, 2003). Invasive spread may exhibit important features such as the presence of a preferential direction for the spread (see Hastings *et al*., 2005 ; Wikle, 2003 ; Morin *et al*., 2009), and depend on of the distance among "suitable" locations. Determining these characteristics is essential to take relevant management decisions.

The Eurasian collared dove (*Streptopelia decaocto*) is one of the most successful invasive birds in North America (Romagosa & Labisky, 2000). This species has been introduced in North America during the early 1980's through Florida. It has been hypothesized that invasion started from Bahamas in the late 1970's, where Eurasian collared-doves escaped from captivity, and after establishing a wild population, reached Florida (Smith, 1987). This species has high colonization capacities. It has invaded Europe in less than 30 years. In fact, in the case of the U.S.A., it took less than 25 years to the European collared dove population to reach the West coast (Dunn & Alderfer, 2006 ; Hooten & Wikle, 2008). Even if this species has not been proven to be a direct threat to other dove species or ecosystems, it is still logically considered as a potential threat since it is an invasive species (Hengeveld, 1993), and therefore might compete with other dove species such as mourning doves (*Zenaida macroura*), white-winged doves (*Zenaida asiatica*) or common ground-doves (*Columbina passerine*) (Romagosa & McEneaney, 2000 ; Romagosa, 2002), or be a disease vector (Romagosa & Labisky, 2000).

Hierarchical modeling has been widely developed and used to estimate site occupancy (Royle & Kéry, 2007). In this framework, hierarchical models are typically based on three components. The first component corresponds to the data, which are the observed quantity. This component is defined conditional on a second component, the state variable (e.g. true

occupancy status); the relationship between these components is accounted for by parameters (e.g. detection probability). A typical recent hierarchical model in population ecology would be constituted by an ecological process underlying occupancy (e.g. balance between extinction and colonization) and a level corresponding to the observation process (which depends on detectability; Royle & Dorazio, 2008). Since non-detection is not equal to absence, this class of models (i.e. including an observation process) is essential when dealing with detection/non-detection data (usually improperly named presence/absence data). Indeed, not accounting for detection issues may lead to substantial bias in population parameters estimation (MacKenzie *et al.*, 2002). This type of data is typically the one available for invasive species, where detectability might be an issue, especially during the beginning of the colonization. Hierarchical models are powerful and flexible and are used to work on a lot of different topics, with many applications to public health and ecology (e.g. Banerjee, 2003 ; Waller & Gotway, 2004 ; Clark, 2007 ; Lawson, 2006 ; Gelman & Hill, 2007 ; Lawson *et al.*, 2008 ; Ntzoufras, 2009) but to our knowledge, they have not been consistently used to assess invasive species dynamics. Importantly, detectability issues (MacKenzie *et al.*, 2006) have seldom been taken into account when addressing hypotheses about vertebrate species distribution or the invasive dynamics of vertebrate species (e.g. Ibarra *et al.*, 2005 ; Ficetola *et al.*, 2007 ; Leprieur *et al.*, 2008). This might be an extremely important issue since ignoring it may lead to underestimation of the actual colonized area and provide erroneous information about the key locations requiring a special regulation effort (e.g. location where invasion is starting and where settlement has not happened yet).

Hierarchical modeling provides convenient means of incorporating biological hypotheses of population dynamics in an explicit way. It is straightforward to express population dynamics parameters (colonization and persistence probabilities) as functions of variables such as density or reproductive success with a hierarchical approach. Nowadays, many

studies use phenomenological models (i.e. models accounting for spatio-temporal patterns of species detection without incorporating specific hypotheses about ecological processes governing species distribution) and emphasize simple descriptions or patterns in data. Such models may lead to satisfactory descriptions of data but are not necessarily easy to interpret biologically. Fewer studies have focused on the development of mechanistic approaches (Bennetts *et al.*, 2001), i.e., models accounting for species presence and detection, with presence expressed according to explicit ecological and biological hypotheses about the dynamics of species distribution. Recent improvements have been made to develop hierarchical models accounting for uncertainty that encompass both spatial and time dimensions (Royle & Kéry, 2007 ; Hooten *et al.*, 2007; Hooten & Wikle, 2008), and directly include scientific insight in model processes (e.g. reaction-diffusion motivation) (Wikle, 2003).

In this paper, we develop a new model based on the one described by Royle & Kéry (2007). In the framework of a Bayesian approach, we extend this model to account more accurately invasive colonization processes, namely, we consider an explicit spatial structure in a dynamic model. This ecological process accounts for the density of occupied sites in the neighborhood. It also considers the influence of distance among sites. Indeed, we expect the occupancy status of close neighboring sites to have a stronger influence on site persistence (or colonization) probability of a site than remote sites. The ecological process also explicitly accounts for the possibility of a directional spread and allows detection of the direction of this spread, if any. To our knowledge, this is the first time that spatial structure is included in a spatio-temporal occupancy model with such an explicit structure formulation. The way the potential anisotropy or directional spread is integrated in the model is a break trough, especially in the framework of occupancy dynamics modeling. We draw a clear and ground-

breaking distinction between first colonization and recolonization since a first colonization might facilitate further colonization events. Our model also includes an observation process, conditional on the ecological one, to deal with detectability issues. Importantly, we consider how previous detections may influence (potentially improve) detectability. Eurasian collared-dove data were collected in the framework of the Breeding Bird Survey (Robbins *et al.* 1986) with the help of volunteer observers. In this case, the presence of an exotic species, easily mistakable with other dove species, might not be properly detected the first time it appears. However, with repeated detection events over consecutive years and the accumulation of external confirmation of this 'unusual' species in the region, observers might improve their ability to detect this species, through a better identification, but more likely because of the knowledge that this species is in fact present in the area. This is the reason why we incorporate a potential 'learning effect' in the observational process, and the resulting gain in detectability.

MATERIAL

European collared dove data are based on the Breeding Bird Survey (B.B.S.; Robbins *et al.* 1986). This program has been monitoring avian populations in North America since 1966. Observers are assigned to a number of routes where they stop 50 times. "Each survey route is 24.5 miles long with stops at 0.5-mile intervals. At each stop, a 3-minute point count is conducted. During the count, every bird seen within a 0.25-mile radius or heard is recorded. […] Over 4100 surveyed routes are located across the continental U.S. and Canada." (B.B.S. website, consulted on 12/09/2008). For each survey route, these raw data include the number of stops where individuals of a given species have been detected, and the total number of individuals detected. Since we are interested in occupancy status (not in abundance) we decided to perform our analyses on the number of stops where the Eurasian Collared-Dove

has been detected. Joseph *et al*. (2006) have shown that abundance methods lead to a larger variation in estimations than detection-absence methods. Moreover, they have shown that presence-absence surveys were more optimal for low budget and low detectability. These points and the fact that detection/non-detection data are often easier information to obtain led us to choose this type of data (but depending on conditions, one can decide to use abundance, like in Wikle, 2003).

We used data from 1986 (first detection of the European Collared-Dove in the USA in the B.B.S. dataset) to 2006. We decided to put a grid on the United States map merging data from all routes contained in the same grid cell. This grid goes from the point of spatial coordinates (24° ; -129°) to the point (57° ; -51°). The side length of a cell is equal to 1°. We did not take into account grid cells that only correspond to Ocean parts. We also did not consider grid cells that do not have at least one neighbor cell, leading to the consideration of a total of 1259 cells.

We define the neighborhood of a cell *i* as the first (N_1) and second (N_2) layers that surround this cell. While cells in the first layer N_1 share a border with cell *i*, cells in the second layer N_2 are separated from cell *i* by one cell, as shown in figure 1. We expect the influence of a site (i.e. cell) occupancy status on another to decrease with increasing distance between the 2 sites. This means that the occupancy status of a close site should have a stronger influence on the probability of occupancy of a given site located further away, like in a diffusive process.

THE MODEL

Occupancy state model

We considered occupancy data obtained by repeating sampling of $i=1,2,\ldots,M$ spatial units (i.e. patches, or "sites", depending on the context, here cells), over $t=1,2,\ldots,T$ periods of time.

Usually, these periods of time refer to significant biological seasons depending on the species of interest. For reference, all parameters used in our model are summarized in Table 1.

The dynamics of the occupancy status will be accounted for by 2 parameters: persistence ϕ (or its complement, local extinction: (1- ϕ)), and colonization γ. Both can be indexed by time and/or site depending on the question of interest. For example, we may consider that site persistence (*i.e.* a cell staying occupied) varies over time depending on the growth rate in a population, *i.e.* stability, growing, or decline. If the population is declining, the persistence of sites will decrease as the number of individuals goes lower. On the other hand, the colonization probability may vary among sites with different characteristics.

Let the occupancy state $Z_{i,t}$ of cell i in year t. If the cell i is occupied at time t, then $z_{i,t}$=1, else, $z_{i,t}$=0. We are interested in the probability of site occupancy $\mu_{i,t}$=Pr($Z_{i,t}$=1|$z_{i,t-1}$) (probability that a cell is occupied conditional on the cell's occupancy state in the previous year). As in population demographic processes of survival and recruitment, here, local extinction and colonization can be used as parameters to model changes in occupancy over time.

Let's define ϕ_t the probability that an occupied site "survives" from time t to $t+1$, that is, given that it was occupied at time t, the probability that it is occupied again at time $t+1$, i.e. ϕ_t=Pr($Z_{i,t+1}$=1 | $z_{i,t}$=1). Where MacKenzie *et al.* (2003) used local extinction probability (1- ϕ_t), we prefer to use its complement, i.e. persistence probability. In addition, let γ_t stand for the local colonization probability, from time t to $t+1$, i.e. γ_t (=Pr($Z_{i,t+1}$=1 | $z_{i,t}$=0). In this model, as in metapopulation systems, local colonization can be viewed as the analog of the recruitment process of individuals in populations. $Z_{i,t}$ is a Bernoulli variable with expected value $\mu_{i,t}$

$$Z_{i,t} \mid z_{i,t-1} \sim Bern(\mu_{i,t}) \quad (1.1)$$

where

$$\begin{aligned} \mu_{i,t} &= \Pr(Z_{i,t} = 1 \mid Z_{i,t-1} = z_{i,t-1}) \\ &= z_{i,t-1}\phi_{t-1} + (1 - z_{i,t-1})\gamma_{t-1} \end{aligned} \quad (1.2)$$

In this model, the occupancy status at time t depends on previous occupancy status. This model can be easily extended. In particular, subsequently, we consider two different stages in the colonization process, and we structure the dynamics parameters to take into account the spatial structure of a point process.

Re-colonization reparametrization

Here, all cells that are not occupied have the same probability of being colonized, i.e. site colonization is random and doesn't depend on the fact that the cell has previously been occupied and then deserted, or has never been colonized at all. However, it is interesting to draw a distinction between two components in the colonization process. We can distinguish between a process of "first colonization" (i.e. the site has never been occupied before, this could be considered as the creation of the site in the dataset), and "recolonization" (i.e. colonization of a site after a previous "extinction"). This is especially relevant for invasive species where initial colonization by some individuals might facilitate further colonization events. For example in the case of invasive plants, seeds can be left by initial colonizers and can germinate long after initial individuals have disappeared, leading to a new and facilitated event of colonization (Harrod & Reichard, 2001 ; Keeley, 2006). While this example is not directly valid for animals, the model proposed in this paper can be applied to plant species with minor modifications and therefore account for this particular situation. Mechanisms for first colonization and recolonization can also be different, e.g., first colonization could be

related to a diffusion type of a process whereas recolonization could be related to the intrinsic dynamics of the local population. In habitat selection theory, one hypothesis is that sites are expected to be chosen upon their quality, and therefore, in this case, "better" sites should be occupied (and reoccupied if abandoned) first, leading to a distinction between first and recolonization probabilities. These two sub-processes of colonization are modeled by splitting the colonization parameter γ_t into a first colonization parameter that will keep the name γ_t, and a recolonization parameter θ_t. We then have to define a new auto-covariate $A_{i,t}$ that will express the availability of a site for first colonization. $A_{i,t}$=1 if the site has never been occupied/colonized before (i.e. for the period T, $A_{i,T}$=1 if $\sum_{t=1}^{T} z_{i,t} = 0$), $A_{i,t}$=0 otherwise.

Therefore, we can formally express $A_{i,t}$ as the indicator function $A_{i,t} = \prod_{k=1}^{t-1}(1-z_{i,k})$, ($k \in \{1,2,...,t-1\}$) as a consequence, sites will have different colonization probabilities depending on whether $A_{i,t}$=1 or $A_{i,t}$=0, respectively, γ_t and θ_t.

Consequently, our model becomes:

$$Z_{i,t+1} \mid z_{i,t} \sim Bern(\mu_{i,t+1}) \qquad (1.3)$$

with

$$\mu_{i,t+1} = \phi_t z_{i,t} + \gamma_t (1-z_{i,t}) A_{i,t} + \theta_t (1-z_{i,t})(1-A_{i,t}) \qquad (1.4)$$

With this model, site colonization can be assessed at two different levels depending on whether it has already been colonized, or not. We have a dynamic model, but we wished to consider the spatial structure by including it in the dynamic parameters.

Spatial structuration

In the model as it stands, the occupancy status of each cell is independent of its environment (i.e., the location of occupied cells and the spatial structure of occupancy).

However, in a large number of cases, we may want to consider the possible impact of the spatial structure of occupancy of several sites (i.e., in a given area) on the fate of individual sites. We may be interested in the influence of variation in climate on a species spatial distribution, or the influence of the spatial structure of the landscape, and the changes that may occur in environmental factors due to human activities. Despite the obvious implications of such processes for both basic and applied ecology, the statistical framework of modeling of spatio-temporal occupancy systems is not well developed, even if some recent efforts have been made, like in Hooten & Wikle (2008) or Zhu *et al*. (2005). Here, we express the spatial structure over the dynamic parameters of persistence, first colonization and recolonization.

Let $\mathcal{N}_i$ represent the set of cells that are neighbors of the cell i, and let n_i be the number of neighbors of cell i (i.e. cardinality of $\mathcal{N}_i$). Then we can define a spatio-temporal autocovariate $D_{i,t}$ as :

$$D_{i,t} = \sum_{j \in N_i} z_{j,t}.w_{ij} \qquad (1.5)$$

w_{ij} is a weight that can be used to specify a difference in the influence of a neighbor site j on the fate of site i according to the locations of sites. It can be based on a simple connection net, where w_{ij}=1 if site j is connected to site i (for example, if the distance between i and j is under the defined/determined threshold "influence" distance), and w_{ij}=0 otherwise (i.e. if site i is not in the influence area of j). We can also set w_{ij} to weight $D_{i,t}$ by the inverse distance of sites j to i, or any other way depending on the decisions made to define the connections network in a relevant manner according to the studied topic. It is important to note that $D_{i,t}$ can be viewed as a surrogate for local density in the vicinity of cell i at time t.

We included this spatio-temporal autocovariate in the model by allowing our dynamics parameters to depend on the autocovariate. Let's take the example of the persistence parameter ϕ. It can simply be expressed as a function of $D_{i,t}$ as follow :

$$\text{logit}(\phi_{i,t}) = a_t + \phi_1 D_{i,t} + \phi_2 D_{i,t}^2 \quad (1.6)$$

It should be noted that ϕ is now indexed by both time period t and site i. We used a quadratic function of $D_{i,t}$ to estimate $\phi_{i,t}$; this function allows ϕ to vary in a way that may lead to a peak at intermediate values of $D_{i,t.}$ Such a pattern, may account for density dependence or "Allee effect" for example (e.g. Courchamp *et al.*, 1999 ; Keyser *et al.*, 2005). Moreover, this formulation permitted us to detect a difference between "random" colonization, as might be expected in a stable metapopulation, and a dynamic of diffusive spread such as might happen in an expanding population. Indeed, a_t will represent the "intrinsic persistence parameter", i.e. the parameter describing what is happening with no neighbor "effect", or when site i doesn't have any occupied neighbor. We decided to allow this "intrinsic persistence parameter" to vary over time, because it can be hypothesized that a_t differs when the population is growing, stable or decreasing. In contrast, ϕ_1 and ϕ_2 represent the influence of occupancy of neighbors on persistence probability, and therefore embodies diffusive or dynamic spread due to gradients in local density or occupancy.

To summarize, our model can be described as follow:

$$Z_{i,t+1} \mid z_{i,t} \sim Bern(\mu_{i,t+1})$$

with

$$\mu_{i,t+1} = \phi_{i,t} z_{i,t} + \gamma_{i,t}(1 - z_{i,t}) A_{i,t} + \theta_{i,t}(1 - z_{i,t})(1 - A_{i,t})$$

where

$$\begin{cases} \text{logit}(\phi_{i,t}) = a_t + \phi_1 D_{i,t} + \phi_2 D_{i,t}^2 \\ \text{logit}(\gamma_{i,t}) = b_t + \gamma_1 D_{i,t} + \gamma_2 D_{i,t}^2 \\ \text{logit}(\theta_{i,t}) = c_t + \theta_1 D_{i,t} + \theta_2 D_{i,t}^2 \end{cases}$$

depending on

$$D_{i,t} = \sum_{j \in N_i} z_{j,t}.w_{ij}$$

The ecological component of our model has now both time and spatial dimensions. While the time dimension is supported by the link between $z_{i,t}$ and $Z_{i,t+1}$, the spatial dimension is encompassed in the weights matrix W (where w_{ij} is the influence of site j on site i). In the following section, we consider how observations are related to the ecological process component of the model and we develop an explicit model for imperfect observation of this process.

Observation Model

One of the principal sources of uncertainty in monitoring data is that due to imperfect detection (or "detectability") of species. That is, a species might be present at some point in space and time yet go undetected. Many modeling approaches and strategies have been devised for dealing with this issue (Williams *et al.*, 2003; MacKenzie *et al.*, 2006).

The simplest way of incorporating detectability (i.e., the probability of contacting an individual of the species given that the species is present in the sites/area sampled) is to consider sampling protocol where a cell i has been visited repeatedly. Then, $K_{i,t}$ defines the number of visits/replications on the cell i at time t, and p the probability that a species/individual is detected during one stop on a road if it is present. $K_{i,t}$ corresponds to the number of routes nested in cell i time the number of stops per route (i.e. 50). $Y_{i,t}$ is the total

number of stops on which a species/individual has been detected in cell *i* at time *t* during the $K_{i,t}$ visits. Then, observation $Y_{i,t}$ follows a binomial with $K_{i,t}$ trials and a probability $p.z_{i,t}$, i.e. :

$$y_{i,t} \sim Bin(K_{i,t}, p.z_{i,t})$$

This means that if the cell is unoccupied, then $Y_{i,t}$ is equal to zero. Otherwise the binomial probability is equal to the detection probability.). $K_{i,t}$ can be viewed as an analogous of the monitoring effort. The higher $K_{i,t}$ is, the higher the global detection probability in the corresponding cell will be. Of course, *p* can be indexed by time and/or site if needed, depending on the question of interest. We have now a full time-space hierarchical model with both ecological and observation processes.

Model adaptations

The model developed previously is a general model that can be easily modified and adapted to a large set of topics, each adaptation having to match the specific questions addressed. Considering specificities of our data, and the questions we are interested in (detection of invasive spread characteristics, improvement of detectability after a first detection), we made some adjustments to the core model previously described.

Spatial structure — One of the main objectives of our work is to provide a model with temporal and a spatial dimensions. To do this, it is important to have a clear and logical definition of the spatial structure. Here, we were interested by the impact of distance between sites on occupancy.

We decided to use the proportion of occupied neighbors in the first and second layers, respectively $D_{1_{i,t}}$ and $D_{2_{i,t}}$ as estimators of local density for the cell *i*.

The proportion of occupied cells in the first layer is just the mean of the number of occupied neighbor sites *j* among the N_1 sites neighboring the cell *i* :

$$D_{1_{i,t}} = \frac{\sum_{j \in N_{1_i}} w_{ij}.z_{j,t}}{N_{1_i}}$$

We defined density in the second layer in the same way:

$$D_{2_{i,t}} = \frac{\sum_{j \in N_{2_i}} w_{ij}.z_{j,t}}{N_{2_i}}$$

Then, local density $D_{i,t}$ of the site i at time t is a weighted sum of relative densities in the first and the second layers.

$$D_{i,t} = \alpha D_{1_{i,t}} + \beta D_{2_{i,t}}$$

Therefore, α and β correspond to the relative contribution of each layer of neighbors. They are estimated during simulations, like dynamic parameters. At this stage in model formulation, the weight w_{ij} is equal to 1 because we consider each site of a layer to be equivalent to the others. According to our hypothesis that close sites should be more influent than distant ones, we expect α to be higher than β.

Anistotropy or d*irectional spread* — Anisotropy is the property of being directionally dependent. Invasion dynamics can be expected to exhibit such a property. The propagation of a species may be governed by a specific environmental gradient (e.g. density, temperature, humidity…) leading to an orientated spread, instead of a simple diffusive expansion. We would like to add this element to our model. Let's consider a spread going from North to South. In this case, the occupancy status of a site *j* located North of a site *i* should be more influent than a site located East or West even more than a site located South.

Let's set a coordinate plane with standard basis, (O, $\vec{x}$, $\vec{y}$) (fig. 2). Let's name δ_1 the angle made by vector $\vec{v}$ (vector indicating the direction of propagation) and abscissa. δ_2 is the angle made vector $\vec{ij}$ (going from site *i*, to site *j*) and abscissa.

As previously exposed, the weights w_{ij} can be used to specify the spatial structure. We used this property to estimate the impact of the direction of propagation on dynamics parameters. We express w_{ij} depending on the direction of the spread and the relative position of site *j* compared to i, which means that we have to express w_{ij} as a function of angle Δ made by vectors $\vec{v}$ and $\overrightarrow{ji}$, as shown in figure 2. It is straightforward to show that $\Delta = \delta_1 - (\pi + \delta_2)$. Then, we find that

$$w_j \propto 1 + \cos(\delta_1 - \pi - \delta_2)$$

and

$$\begin{aligned} w_j &\propto 1 + \cos(\delta_1 - \pi - \delta_2) \\ &\propto 1 - \cos(\delta_1)\cos(\delta_2) - \sin(\delta_1)\sin(\delta_2) \end{aligned} \qquad (9)$$

δ_2 is calculated from coordinates of sites *i* and *j* (respectively $\begin{pmatrix} x_i \\ y_i \end{pmatrix}$ and $\begin{pmatrix} x_j \\ y_j \end{pmatrix}$). We have :

$$\cos(\delta_2) = \frac{x_j - x_i}{\sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2}}$$

and

$$\sin(\delta_2) = \frac{y_j - y_i}{\sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2}}$$

Therefore, the only unknown parameter left is δ_1 (indicating the angle of propagation) and it will be estimated during simulations.

We have now a hierarchical time-space model, with an ecological process accounting for the impact of distance between sites through α and β, and the direction of a spread and the relative position of site *j* to *i*, through δ_1.

Observation process — Concerning the observation process, we would like to deal with a problem that may occur due to the way raw data are collected in the B.B.S. These data are collected by volunteers and are subject to various types of uncertainty, including observer error (Sauer *et al.*, 1994). We think that it is interesting to modify the observation process because of possible misidentifications between European collared dove and for example ringed turtle dove, and the propensity of observer to consider unlikely the presence of European collared dove because it is an invasive species (i.e. non endemic species) In fact, we think that a previous confirmed observation may lead to larger subsequent detection probability. In the framework of invasive species and volunteers, the underlying idea is that persons collecting data may not identify a new species that is not supposed to be there (non indigenous), and easily confound it with another. However, as soon as this species has been detected without doubt, detection skills of observers may improve, and they may no longer consider the presence of this exotic species unlikely. We note that this is analogous to a "behavioral response" in classical capture-recapture modeling (Otis *et al.*, 1978 ; or Williams *et al.*, 2002, chapter 14).

We consider the previous observation process:

$$y_{i,t} \sim Bin(K_{i,t}, p_{i,t}.z_{i,t})$$

but this time we allow detection probability *p* to vary with time and site. We can accommodate this behavioral response by including an effect on detection probability of prior detection of the species:

$$\text{logit}(p_{i,t}) = \lambda_1 + \lambda_2(1 - A'_{i,t})$$

Here $A'_{i,t}$ corresponds to the availability for a first detection. If the species has never been previously detected in patch *i* before time *t*, then $A'_{i,t} = 1$. Else, $A'_{i,t} = 0$. Then λ_1 is the probability of first detection (the species has never been detected before), and λ_2 the gain in detectability due to a previous detection.

Our model now incorporates an observation process that considers improvement in detectability. Consequently, we can estimate a part of the error due to observer inexperience.

Bayesian analysis and implementation in WinBUGS

Because of the conditional specification of this time-space dynamic site occupancy model, Bayesian analysis appears well suited, via Markov chain Monte Carlo. The model was estimated using Gibbs sampling (Casella & George, 1992) which is based on drawing samples of each unknown quantity from their "full-conditional" distributions, i.e., the distribution of a parameter conditional on all other unknown quantities and the data (Royle & Kéry, 2007).

Implementing our model with a software program such as WinBUGS is straightforward. For the analysis presented in this publication, models were implemented under the free software package WinBUGS 1.4.3 (Lunn *et al.*, 2000) called from R 2.9.0 (R Development Core Team, 2008) using the R add-on library R2WinBUGS (Sturtz *et al.*, 2005). We ran 3 chains and based our inference on 100.000 samples from the posterior distribution of parameters, after 20.000 discarded iterations. The code for this model is provided in appendix A. All results are presented with standard deviation.

To validate our model, we compared probabilities of detecting at least one individual in a cell for each year from 1986 to 2005 to detection data (i.e. the corresponding cell has been detected or not to be occupied) of the following year (i.e. from 1987 to 2006) using the R package "*ROCR*" by Sing *et al.* (2009). We provide the result of the Area Under the Curve for the corresponding ROC diagrams for the last 15 years.

RESULTS

With a mean AUC for the last 15 years equal to 0.72 ± 0.14, reaching 0.76 ± 0.07 for the last 8 years, our model can be considered to be a fair estimation of the invasion process.

Spatial structure

Distance – With respective values of α=0.79 ±0.12 and β=0.87 ±0.09, weights for layers 1 and 2 do not reveal a clear prevalence of a layer on the other. Surprisingly, and in opposition with our predictions, this means that we failed to detect a clear influence of distance on dynamics parameters, at least at the scale considered here.

Invasion spread direction – We detected a preferential direction for the invasion spread with δ_1=140.3° ±14.91. This approximately corresponds to a spread going from South East to North West. This corresponds to empirical observations that indicate a first phase of slow colonization towards North, followed by a second phase of rapid expansion towards West as shown in figure 3. The invasion progresses mainly from East to West, with a component South to North. Starting in Florida, this invasion, to proceed, could not have been done other than going towards North. After that, a limitation due to latitude, with temperatures not being suitable for this species over the North limit of the invasion front is likely to appear.

On figure 3, we provide an example that permits to compare our estimations for sites that have a high probability of being colonized (hatched area, p>0.5) in 2006 based on 2005 data, and actual occupancy observed in 2006 (gray cells). We see that estimations are accurate and globally correspond to the actual observed occupancy. This type of prediction map could be used for management decision to contain the propagation of an invasive species. Here, using the estimation of spread direction and the impact of density, we illustrate the Western invasion process front and the likely recolonization of some abandoned sites.

Density – Site persistence probability (black line, fig. 4) increases at low density and remains high at intermediate densities before dropping at the highest densities. First colonization probability (dashed line, fig.4) is globally low but exhibits a quadratic relationship with density, with virtually no first colonization at low and high density, and a slightly higher value at a short range of medium densities. Recolonization probability (dotted line, fig.4) is quite high at low to intermediate densities, and then decreases at higher densities.

Detectability

Initial detection probability estimate at the stop level is equal to 0.03 (λ_1= -3.44 ±0.28). Surprisingly, after a previous detection, detection probability drops around 0.01 (λ_2=-1.633 ±0.5395). With such a low detection probability, not accounting for this issue would have led to underestimation of occupancy probabilities and therefore the area occupied.

DISCUSSION

We have developed a time-space hierarchical model accounting for an invasion process via estimation of site persistence probability, first colonization and recolonization, and their relationship with density. We were able to detect the direction of an invasive spread, but we have not detected any influence of distance between occupied sites on dynamics parameters. Detection probability estimates were low and decreased after the first detection of the presence of the species in a site, which is counterintuitive.

Our model provides an estimate of the direction of spread direction consistent with empirical observations: globally, the spread direction is toward the northwest. The invasion started in the early 1980's through Florida. Colonization really started in direction of the North in the 1990's, before expanding towards West in the late 1990's. If the spread had not followed a specific direction, and had been globally the same in every spatial direction, the

estimate for standard error for δ_1 would have been large. Compared to Wikle (2003), the spatio-temporal model we propose accounts for the possibility of a directional diffusion. Moreover, here the spatial structure is explicitly specified, which may be interesting especially if structures are known (or expected) to facilitate or hamper population/individual movements. The explicit formulation of the spatial structure may be used to reflect a particular underlying geographical structure that may affect the spread of the invasion. In this study, we have considered the simplest case with no environmental or topographical barriers, but this could be supported by the weight matrix. If we had considered these elements, estimation of the direction of the spread, and impact of the distance might have been different. For example, we might expect the Great Plains to facilitate the invasion process, while the Rocky Mountains might stop or slow down this propagation. On the other hand, human activities such as farming, or even bird feeding in cities might help the collared dove to settle in areas that would not be particularly suitable due to lower temperature in the North.

A development of this model would be to estimate the spread direction for each year to have a more accurate and detailed description of the colonization process and of the local anisotropic conditions, but this would require tremendous computing time and calculus capacities. This year-specific approach may permit us to evaluate if propagation of the invasive is facilitated in the South once the species has reach the Mexican border because of the suitable climatic conditions. Moody & Mack (1988) and Hajek *et al.* (1996) have discussed the importance of targeting control efforts at the leading edge of invading populations, especially at recently colonized sites ahead of the main body of the invasion. If these sites are colonized, this spread can speed the overall invasion progress. Accurate estimation of the spread direction and colonization probabilities can greatly improve how and where control efforts should be devoted. For this purpose, estimates provided by the type of model we developed permit the construction of predictive maps for the invasion progress of a

particular species, which should help target management efforts to locations where they would be more useful and pertinent.

The relationship between site persistence probability and density is quadrating, which may reflect an Allee effect at low densities. The high persistence probability at intermediate to high densities indicates that once the species has started to spread and has settled in a location, it tends to stay there, which leads to a durable invasion. The slight observed decrease at the highest densities might be due to extrapolation beyond the observed densities. The relatively high recolonization probability at low density indicates that even if a site is abandoned, it will be re-occupied. The decrease of recolonization at higher densities is probably due to the fact that, there is no available site and they all stay occupied, considering the high persistence probability at these densities. The low first colonization probability at low densities indicates that there is no first colonization event in isolated areas. The higher first colonization probability at intermediate density shows that colonization events occur at the leading edge of the invading population (i.e. margin of geographical distribution). The distinction between first colonization and recolonization may be useful to deal with a wide range of topics where these two processes (or "subprocesses") are influenced by different factors. For example, in the framework of the selection of a breeding site, we may expect higher-quality sites to be chosen first (Møller, 1982 ; Newton & Marquiss, 1982), and when they get deserted for any reason by individuals (e.g. death, dispersal,…), we may expect these abandoned sites to be reoccupied first, before new breeding sites are “created”. Here, we have seen that first colonization and recolonization are influenced differently by density. For management purposes, it may be useful to distinguish these two subprocesses since they might not be equally affected by control measures, the latter may even lead to opposite effects in each subprocess, leading to an unwanted response.

We expected that weights attributed to site occupancy status decrease as the distance between sites increases. It is logical to think that sites close to already occupied sites have higher persistence and colonization probabilities than distant sites (especially in the case of an avian invasive species). For example, the proximity of a source of dispersing individuals has been shown to increase the likelihood of a given site being colonized (Kolar & Lodge, 2001 ; Lockwood *et al.*, 2005). However, our results do not corroborate this hypothesis: we have not detected any relationship between distance among sites at the scale considered and dynamics parameters. The most likely explanation for this non-detection of a distance effect is the scale we are working at (i.e. the size of the cell). At a smaller scale, this effect might have been detected. This may depend on the speed of the invasion: indeed, we might expect that as the invasion speed increases, the site influence area increases in size. In other words, if a species colonizes an area rapidly, distant sites are as likely to be quickly colonized as close sites. At a small scale, variation in distance might not allow detection of an impact of the distance since neighboring sites might all have a high colonization probability if the species is mobile. On the other hand, if the scale is too large, all sites will have a low colonization probability no matter their distance from an occupied cell, because even directly neighboring cells might be over the influence area of the occupied point within the cell. Therefore we could not detect an impact of distance on colonization probabilities if the scale we are working at does not allow us to draw a distinction between significant classes of distance (for the colonization process). Especially in the case of avian invasive species, individuals can be expected to have high dispersive capacities. Therefore, the distance effect would be small, and hard to detect. It would be interesting to try to re-run this model using a grid including a larger number of cells (i.e. reduce the size of cells) and a more important number of layers to

incorporate more distance classes, but once again, such an advanced decomposition requires an important amount of data and tremendous computing capacities.

While detection probability estimates can seem really low, it should be noted that it corresponds to the detection probability at the stop level, therefore global detection probability for the cell is significantly higher. A simple calculation shows that when the species is present in a cell, the probability of detecting at least once the species is $1-(1-p)^K$. For example, even if we have p=0.02, the probability of detecting at least once the species in a cell with only one route (i.e. K=50, since there are 50 stops per route) will be $1-0.98^{50}=0.64$. This probability jumps to 0.87 if there are 2 routes in the cell (K=100), and increases up to 0.95 for 3 routes. Furthermore, the deterioration of detectability after a first detection is probably due to the type of data we used and the scale considered. We merged B.B.S. occupancy data from several routes when variability among routes can sometimes be substantial. Consequently, there can be a very large intrinsic variance in patterns of detections within a cell. The artificial homogenization resulting from the merging of several routes into a cell may diminish our ability to detect sensitive effects in the data. We think that the combination of the scale considered and the grouping of data from several routes within cells might have led to a poor estimation of detection probability. Analyses at a smaller scale with abundance data may provide more reliable detectability estimates and may allow assessment of the "learning effect" due to a previous detection, if any. Additionally, the B.B.S. data are subject to multiple sources of uncertainty, especially within-site variability (Link *et al.*, 1994; Sauer *et al.*, 1994), which should be considered seriously if such data were used to fit abundance models.

The model developed in this paper is based on a simple, and yet powerful (in terms of versatility and quality of estimations), hierarchical formulation aimed at assessing the

occupancy status of sites as functions of persistence and colonization probabilities. The basic model has been described and used in MacKenzie *et al.* (2003) (who gave the basic likelihood formulation) and Royle & Kéry (2007) (who provided the hierarchical Bayesian form) Conditionally on the presence of the species/individuals in a site, data (i.e. observations) are then expressed as a function of detection probability. It is important to realize that this class of models can be fitted at very different ecological scales. In fact, as described in Royle & Dorazio (2008), if you consider a 'static' system, the ecological scale (e.g. population of individuals, metapopulations) is express by a size parameter (e.g. number of individuals/species). Dynamic systems include parameters such as survival and recruitment probabilities for populations, analogous to extinction and colonization parameters in metapopulations and communities systems. Therefore, such models can be used to address a large variety of topics. The model shown in this paper has been developed to be fitted at scales as different as U.S.A. colonization by the Eurasian Collared-dove, or nest selection inside a cliff by black-legged kittiwake *Rissa tridactyla*.

Our model presents some components specific to invasive species that can easily be modified. For example, the quadratic expression of dynamics parameters as a function of density allows us to account for peaks, and therefore, is useful to consider a priori biological assumptions. However, this is an arbitrary formulation. It is possible to develop an individual-based model to account for a more accurate relationship between dynamics parameters and density (in terms of biological and/or behavioral mechanisms). Moreover, in the current state of the model, ϕ_1 and ϕ_2 do not depend on time. We made this decision because we assumed that the way the neighborhood influences a site's fate does not depend on the growth status of the population. This assumption can be easily relaxed. Moreover, we modeled $\gamma_{i,t}$ and $\theta_{i,t}$ as $\phi_{i,t}$, but this can be changed depending on whether investigators want to have all the dynamics parameters to be influenced by the spatial structure of occupancy, or

not. The spatial spread of an invasive species follows successful establishment, which is a step driven by colonization and extinction processes (Eraud *et al.*, 2007). Several factors associated with variation in colonization probabilities were highlighted by recent theoretical work on invasion and metapopulation theory. Environmental heterogeneity and distance between patches of suitable habitats (Hastings *et al.*,2005) for example have an influence on persistence and colonization probabilities (With, 2002) and can be modeled differently for each dynamics parameter.

Modeling invasive species population dynamics is often based on population density and growth rate estimation (Veit & Lewis, 1996 ; Taylor & Hastings, 2004 ; Le Maitre *et al.*, 2008), and efficient ways of estimating occupancy are emerging via the development of hierarchical models. However, these models are not yet extensively used to deal with invasive species. Several mathematical formulations have been developed to define, estimate and ultimately model the spatial expansion of spreading population (e.g. Van den Bosch *et al.*, 1992; Neubert & Caswell, 2000 ; Ortega-Cejas *et al.*, 2004). But these classical models, despite the interesting advances they represent, suffer some limitations in terms of assessment and understanding of the expansion process (Eraud *et al.*, 2007) (e.g. lack of estimates of relevant components such as occupancy rate, local colonization probability). Furthermore, existing models are usually fitted to distribution maps derived from observational counts of organisms, assuming perfect detectability of individuals or species (Eraud *et al.*, 2007). Models such as the one developed in this paper, through their high modularity and consideration of detectability issues, can expand and complete observations given by the above classical models.

Invasive species have become a major problem not only in North America but throughout the world (Cronk & Fuller, 1995 ; Williamson, 1996 ; Williamson, 1999 ; Mooney & Cleland, 2001 ; Molnar *et al.*, 2008). Investigations of the spatio-temporal dynamics of invasive species will permit predictions about future spread as well as about the likely efficiency of management actions designed to control such spread (e.g., Wikle, 2003). Ultimately, management and control of invasive species has to be conducted in an integrative framework where ecological, statistical and dynamical approaches as to be coupled with genetic studies. As a matter of fact, because of possible hybridization with sympatric species, it is necessary to couple occupancy observations with potential genetic consequences of this invasion on other close species to fully realize the impact of invasive species on local ecosystems.

The main aim of this work was to develop a hierarchical model that encompasses time and space dimensions in a convenient and flexible way. Moreover, the distinct responses to density variation of first and re-colonization probabilities confirm that it is essential to consider these two sub-processes to understand the global colonization process to fully understand how an invasion progresses, and ultimately be able to make suitable management decisions.

ACKNOWLEDGEMENTS

The authors thank Marc Kéry, Beni Schmidt, Robert Dorazio, Beth Gardner, Adam Green. The authors are also grateful to the two reviewers for their helpful comments.

LITERATURE CITED

Andow D.A., Kareiva P.M., Levin S.A. & Okubo A., 1990. Spread of invading organisms. *Landscape Ecology*. 4: 177—188.

B.B.S. website (consulted on 12/09/2008). http://www.pwrc.usgs.gov/bbsapps/

Banerjee S., 2003. Hierarchical Modeling and Analysis for Spatial Data. *Chapman & Hall/CRC*. 472p.

Bennett R.E., Nichols J.D., Lebreton J.D., Pradel R., Hines J.E. & Kitchen W.M., 2001. Methods for estimating dispersal probabilities and related parameters using marked animals. In : *Dispersal* (Clobert J., Danchin E., Dhondt A.A. & Nichols J.D., eds). Oxford University Press. 3—17.

Casella G. & George E.I., 1992. Explaining the Gibbs Sampler. *The American Statistician*. 46: 167—174.

Clark J.S., 2007. Models for Ecological Data: An Introduction. *Princeton University Press*. 632p.

Courchamp F., Clutton-Brock T. and Grenfell B., 1999. Inverse density dependence and the Allee effect. *Trends in Ecology & Evolution*. 14: 405—410.

Cronk Q.C.B. & Fuller J.L., 1995. Plant Invaders. The Threat to Natural Ecosystems. *London: Chapman and Hall*. 241p.

DAISIE European Invasive Alien Species Gateway (http://www.europe-aliens.org/)

Dunn J.L. & Alderfer J., 2006. National Geographic Field Guide to the Birds of North America, Fifth Edition. *National Geographic Society, Washington*. 504p.

Eraud C., Boutin J.-M., Roux D. & Faivre B., 2007. Spatial dynamics of an invasive bird species assessed using robust design occupancy analysis: the case of the Eurasian collared dove (*Streptopelia decaocto*) in France. *Journal of Biogeography*. 34: 1077—1086.

Ficetola F.G., Coïc C., Detaint M., Berroneau M., Lorvelec O. & Miaud C., 2007. Pattern of distribution of the American bullfrog *Rana catesbeiana* in Europe. *Biological Invasions*. 9: 767—772.

Gelman A. & Hill J., 2007. Data Analysis using regression and multilevel/hierarchical models. *Cambridge University Press, Cambridge*. 625p.

Glowka L., Burhenne-Guilmin F. & Synge H., 1994. A guide to the convention on biological diversity. *IUCN, Gland*. 173p.

Gurevitch J. & Padilla D.K., 2004. Are invasive species a major cause of extinctions? *Trends in Ecology & Evolution*. 19: 470—474.

Hajek, A.E., Elkinton, J.S. & Witcosky, J.J., 1996. Introduction and spread of the fungal pathogen *Entomophaga maimaga* (*Zygomycetes: Entomophthorales*) along the leading edge of gypsy moth (*Lepidoptera: Lymantriidae*) spread. *Environmental Entomology*. 25: 1235—1247.

Harrod R. J. & Reichard S., 2001. Fire and invasive species within the temperate and boreal coniferous forests of western North America. 95—101 *in* K. E.M. Galley and T. P.Wilson, editors. *Proceedings of the invasive species workshop: the role of fire in the control and spread of invasive species*. Miscellaneous publication 11. Tall Timbers Research Station, Tallahassee, Florida

Hastings A., Cuddington K., Davies K.F., Dugaw C.J., Elmendorf S., Freestone A., Harrison S., Holland M., Lambrinos J., Malvadkar U., Melbourne B.A., Moore K., Caz Taylor C. &

Thomson D., 2005. The spatial spread of invasions: new developments in theory and evidence. *Ecology Letters*. 8: 91—101.

Hengeveld R., 1993. What to do about the north American invasion by the collared dove? *Journal of Field Ornithology*. 64: 477—489.

Hooten M.B., Wikle C.K., Dorazio R.M. & Royle J.A., 2007. Hierarchical spatiotemporal matrix models for characterizing invasions. *Biometrics*. 63: 558—567.

Hooten M.B. & Wikle C.K., 2008. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environmental and Ecological Statistics*. 15: 59—70.

Ibarra A.A., Park Y.S., Brosse S., Reyjol Y., Lim P. & Lek S., 2005. Nested patterns of spatial diversity revealed for fish assemblages in west European river. *Ecology of Freshwater Fish*. 14: 233—242.

Joseph L.N., Field S.A., Wilcox C. & Possingham H.P., 2006. Presence–Absence versus Abundance data for monitoring threatened species. *Conservation Biology*. 20: 1679—1687.

Keeley J.E.,2006. Fire Management Impacts on Invasive Plants in the Western United States. *Conservation Biology*. 20: 375—384.s

Keyser P.D., Guynn Jr D.C. & Hill Jr. H.S., 2005. Density-dependent recruitment patterns in white-tailed deer. *Wildlife Society Bulletin*. 33: 222—232.

Kolar C.S. & Lodge D.M., 2001. Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution*. 16: 199—204.

Kot M., Lewis M.A. & Van Den Driessche P., 1996. Dispersal data and the spread of invading organisms. *Ecology*. 77: 2027—2042

Lawson A.B., 2006. Statistical Methods in Spatial Epidemiology. *Wiley Series in Probability and Statistics*. 424p.

Lawson A.B., 2008. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. *Chapman & Hall/CRC*. 368p.

Le Maitre D.C., Krug R.M., Hoffmann J.H., Gordon A.J. & Mgidi T.N., 2008. *Hakea sericea*: Development of a model of the impacts of biological control on population dynamics and rates of spread of an invasive species. *Ecological Modelling*. 212: 342—358.

Leprieur F., Beauchard O., Blanchet S., Oberdorff T. & Brosse S., 2008. Fish invasions in the world's river systems: when natural processes are blurred by human activities. *Public Library of Science, Biology*. 6: e28.

Link W.A., Barker R.J. & Sauer J.R., 1994. Within-site variability in surveys of Wildlife populations. *Ecology*. 75: 1097—1108.

Lockwood J.L. , Cassey P., Blackburn T., 2005. The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution*. 20: 223—228.

Lunn D.J., Thomas A., Best N. & Spiegelhalter D., 2000. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10: 325—337.

Mack R. N., 1985. *in* Studies in Plant Demography, ed. White, J. *Academic, London*. 127—142.

MacKenzie D.I., Nichols J.D., Hines J.E., Knutson M.G. & Franklin A.D., 2003. Estimating site occupancy colonization and local extinction when a species is detected imperfectly. *Ecology*. 84: 2200—2207.

MacKenzie D.I., Nichols J.D., Lachman G.B., Droege S., Royle A.J. & Langtimm A.C., 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*. 83: 2248—2255.

MacKenzie D.I., Nichols J.D., Royle J.A., Pollock K.H., Bailey L.L. & Hines J.E., 2006. *Occupancy Estimation and Modeling*. Academic Press. San Diego, California, USA. 324p.

Møller A.P., 1982. Characteristics of Magpie Pica pica territories of varying duration. *Ornis Scandinavica*. 13: 94—100.

Molnar J.L., Gamboa R.L., Revenga C. & Spalding M.D., 2008. Assessing the global threat of invasive species to marine biodiversity. Frontiers in Ecology and the Environment. 6: 485—492.

Moody M.E. & Mack R.N., 1988. Controlling the spread of plant invasions, the importance of nascent foci. *Journal of Applied Ecology*. 25: 1009—1021.

Mooney H.A. & Cleland E.E., 2001. The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences*. 98: 5446—5451.

Morin R.S., Liebhold A.M. & Gottschalk K.W., 2009. Anisotropic spread of hemlock woolly adelgid in the eastern United States. *Biological Invasions*. 11: 2341—2350.

Neubert M.G. & Caswell H., 2000. Demography and dispersal: calculation and sensitivity analysis of invasion speed for structured populations. *Ecology*. 81: 1613—1628.

Newton M. & Marquiss I., 1982. Habitat preference in male and female sparrowhawks *Accipiter nisus*. *Ibis*. 124: 324—328.

Ntzoufras, I. 2009, Bayesian Modeling using WinBUGS. John Wiley & Sons, Inc. 489p.

Okubo A., 1980. Diffusion and ecological problems: mathematical models. *Springer-Verlag, Berlin, Germany*.

Ortega-Cejas V., Fort J. & Méndez V., 2004. The role of the delay time in the modeling of biological range expansions. *Ecology*. 85: 258—264.

Otis D.L., Burnham K.P., White G.C. & Anderson D.R., 1978. Statistical inference from capture data on closed animal populations. *Wildlife Monographs*. 62: 1—135.

R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Robbins C.S., Bystrak D. & Geissler P.H., 1986. The Breeding Bird Survey: Its First Fifteen Years, 1965-1979. *U.S. Dept. of the Interior, Fish and Wildlife Service, Washington, D.C. (USA)*. 205p.

Romagosa C.M. & Labisky R.F., 2000. Establishment and Dispersal of the Eurasian Collared-Dove in Florida. *Journal of Field Ornithology*. 71: 159—166.

Romagosa C.M., 2002. Eurasian Collared-Dove (*Streptopelia decaocto*). In The Birds of North America, No. 630 (A. Poole and F. Gill, eds.). *Birds of North America, Inc., Philadelphia, PA.*

Romagosa, C.M. & McEneaney T., 2000. The Eurasian Collared-Dove in North America and the Caribbean. *North American Birds*. 53: 348—353.

Royle J.A. & Dorazio R.M., 2008. Hierarchical Modeling and inference in ecology (The analysis of data from populations, metapopulations, and communities). *Academic Press*. 464p.

Royle J.A. & Kéry M., 2007. A Bayesian state-space formulation of dynamic occupancy models. *Ecology*. 88: 1813—1823.

Sauer J. R., Peterjohn B.G. & Link W.A., 1994. Observer differences in the North American Breeding Bird Survey. *Auk*. 111: 50—62.

Sing T., Sander O., Beerenwinkel N. & Lengauer T., 2009. ROCR: Visualizing the performance of scoring classifiers. *R package version 1.0-4*. http://CRAN.R-project.org/package=ROCR

Skellam J.G., 1951. Random dispersal in theoretical populations. *Biometrika*. 38: 196—218.

Smith P. W. 1987. The Eurasian Collared-Dove arrives in the Americas. *American Birds*. 41: 1370—1379.

Sturtz S., Ligges U. & Gelman A., 2005. R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software*. 12: 1—16.

Taylor C.M. & Hastings A., 2004. Finding optimal control strategies for invasive species: a density-structured model for *Spartina alterniflora. Journal of Applied Ecology*. 41: 1049—1057.

Van den Bosch F., Hengeveld R. & Metz J.A.J., 1992. Analysing the Velocity of Animal Range Expansion. *Journal of Biogeography*. 19: 135—150.

Veit R.R. & Lewis M.A., 1996. Dispersal, Population Growth, and the Allee Effect: Dynamics of the House Finch Invasion of Eastern North America. *The American Naturalist*. 148: 255—274.

Waller L.A & Gotway C.A., 2004. Applied Spatial Statistics for Public Health Data. *Wiley Series in Probability and Statistics*. 520p.

Wilcove D.S., Rothstein D., Dubow J., Phillips A. & Losos E., 1998. Quantifying threats to imperiled species in the United States. *Bioscience*. 48: 607—615.

Wikle C.K., 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*. 84: 1382—1394.

Williams B.K., Nichols J.D. & Conroy M.J., 2002. Analysis and Management of Animal Populations. *San Diego, California, Academic Press*.

Williamson M., 1996. Biological Invasions. *Chapman and Hall, London*. 244p.

Williamson M., 1999. Invasions. *Ecography*. 22: 5—12.

Zhu, H., Huang H.-C. & Wu C.-T., 2005. Modeling spatial-temporal binary data using Markov random fields. *Journal of Agricultural, Biological, and Environmental Statistics*. 10: 212—225.

Tables and figures

Table 1.

Table 1. Parameters used and estimated in the model

Parameter	Description	Algorithm notation
$Z_{i,t}$	Occupancy state of cell i during year t.	`z[i,t-1]`
$\mu_{i,t}$	Site occupancy probability	`muZ[i,t]`
$\phi_{i,t}$	Persistence parameter	`phi[i,t-1]`
$\gamma_{i,t}$	First colonization parameter	`gamma[i,t-1]`
$\theta_{i,t}$	Recolonization parameter	`theta[i,t-1]`
$A_{i,t}$	Availability of a site for first colonization	`A[i,t-1]`
$D_{i,t}$	Local density	`D[i,t-1]`
w_{ij}	Weight reflecting importance of occupancy status of cell *j* on future occupancy status of site *i*	*(computed in the algorithm)*
a_t, b_t, c_t	Respectively intrinsic persistence, first colonization and recolonization parameters	`lphi0[t]`, `lgamma0[t]`, `ltheta0[t]`
ϕ_1, γ_1, θ_1	Respectively linear factors of persistence, first colonization and recolonization parameters to local density	`lphi1`, `lgamma1`, `ltheta1`
ϕ_2, γ_2, θ_2	Respectively quadratic factors of persistence, first colonization and recolonization parameters to local density	`lphi2`, `lgamma2`, `ltheta2`
$K_{i,t}$	Number of replications on the cell *i* at time *t*	`K[i,t]`
p	Detection probability	`p[i,t]`
a, β	Respectively relative contributions of layers 1 and 2 density to local density	`alpha, beta`
δ_1	Direction of propagation	`dirSpread`
$A'_{i,t}$	Availability for a first detection	`Aprim[i,t]`
λ_1	First detection probability	`ObsInit`
λ_2	Gain in detectability after a previous detection	`DeltaObs`

Figure 1. Representation of the 2 layers of neighboring cells, used in our model in estimation of local density. *In black, patch 'i', in grey first layer N_1 of cells, hatched: second layer N_2 of neighboring cells*

Figure 2. Diagrammatic representation of points and vectors used in anisotropy modeling. *δ_1 is the angle made by the vector of invasion propagation $\vec{v}$ and abscissa. δ_2 is the angle made vector $\vec{ij}$ (going from site i, to site j) and abscissa.*

Figure 3. Eurasian Collared-Dove site occupancy for 1996 (*black area*) and 2006 (*grey area*), and the corresponding estimations of probability of being colonized in 2006 (*hatched area*, p>0.5).

Figure 4. Estimates of Dynamics parameters as a function of local density *D. Black line is site persistence probability φ. Dashed line is first colonization probability γ. Dotted line is recolonization probability θ.*

Figure 1.

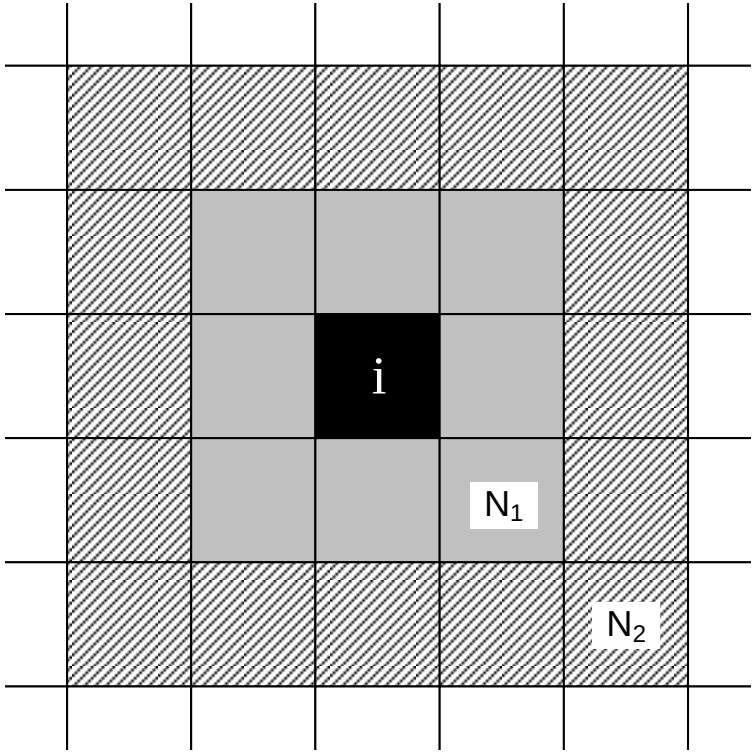


Figure 2.

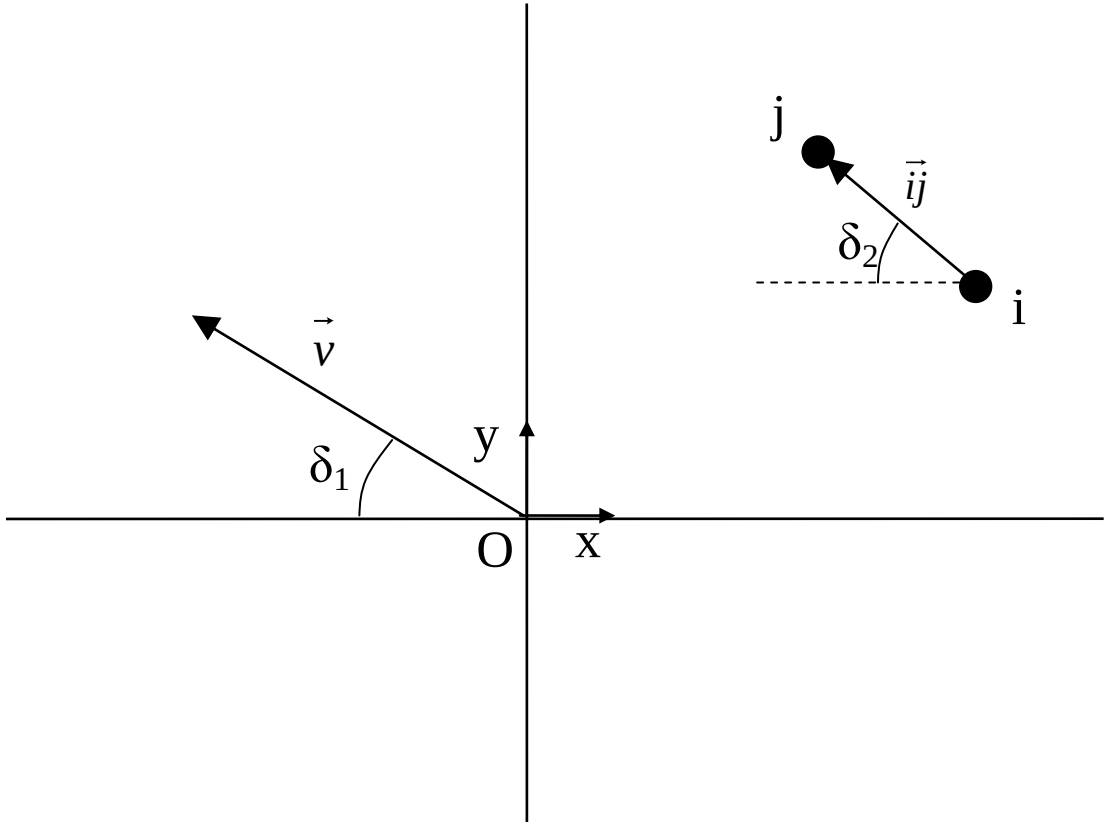


Figure 3.

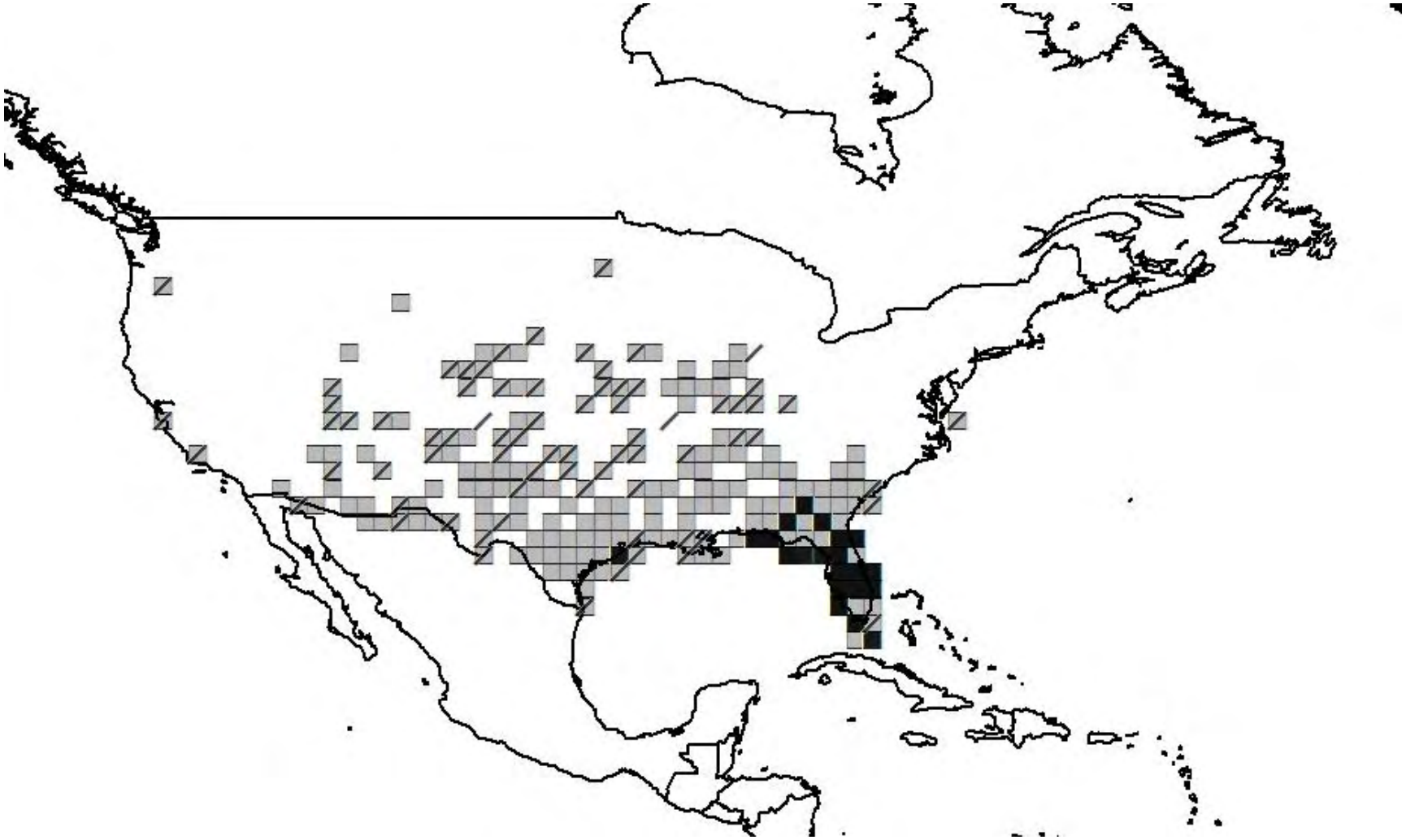
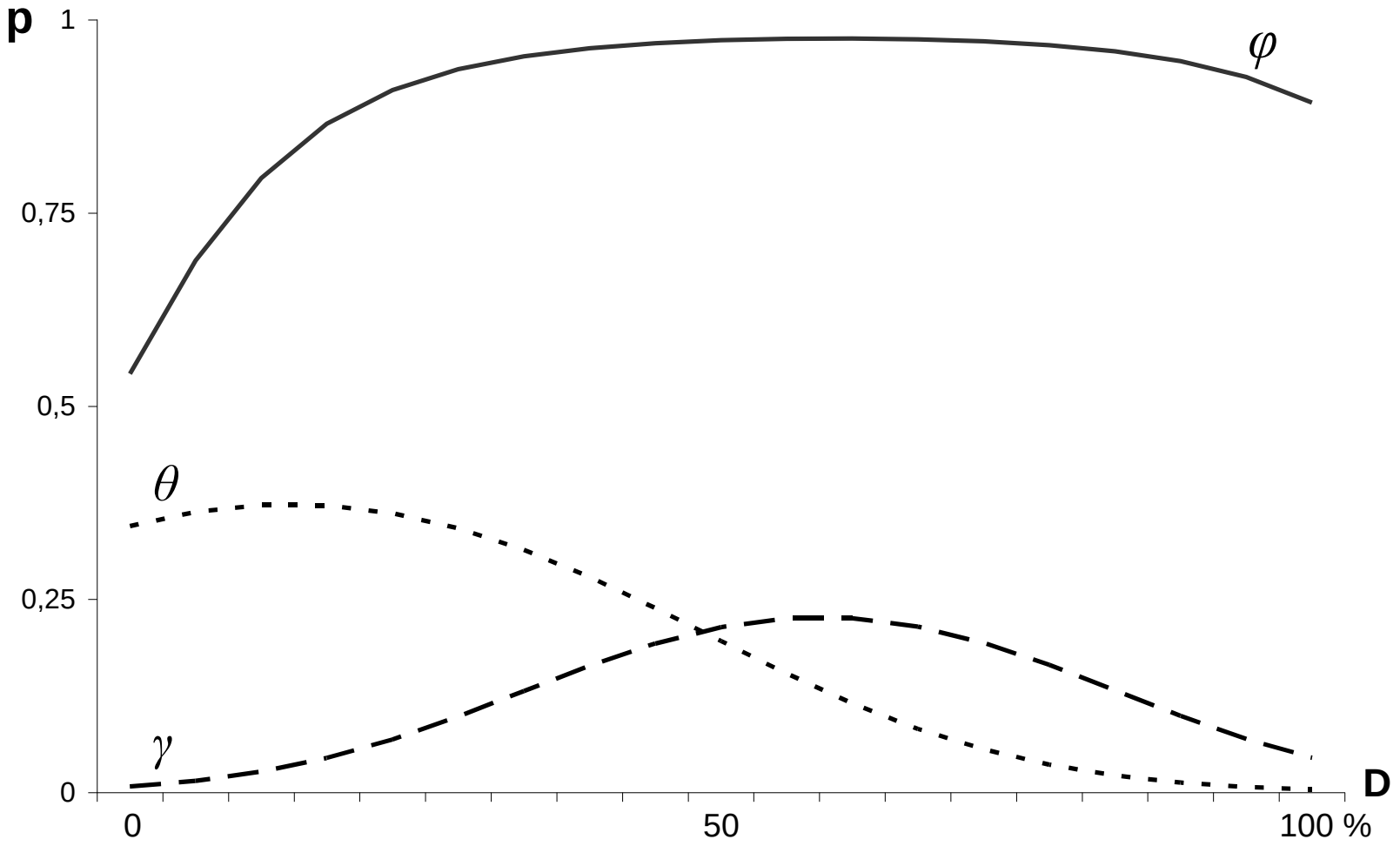


Figure 4.



CHAPTER III

1

Assessing hypotheses about nesting site occupancy dynamics

Bled F.[1]*, Royle J.A.[2] & Cam E.[1]

[1] Laboratoire Évolution et Diversité Biologique, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 9, France.

[2] U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland 20708, USA.

(Ecology, in press)

***** Corresponding author (bled@cict.fr)

Abstract

Understanding the distribution of animal populations has long been central to population ecology. For mobile species living in a heterogeneous environment, evolutionary ecology provides a conceptual framework to address the relationship between individual decisions and habitat features. Hypotheses about habitat selection assume that individuals select breeding habitat based on expected fitness in different habitat, on condition that *i)* competition does not prevent site choice, *ii)* individuals are free to move, *iii)* individuals have perfect knowledge of potential breeding sites. In addition to direct assessment of habitat quality through physical or biotic criteria, the relationship between habitat quality and fitness may be partly reflected by breeding success of individuals, which may in turn be used to assess habitat quality. Habitat quality may also be assessed through local density: if high-quality sites are preferentially used, high density may reflect high-quality habitat.

Here we assessed whether site occupancy dynamics vary with site surrogates for habitat quality, We modeled nest site use probability in a seabird subcolony (the kittiwake *Rissa tridactyla*), over a 20 year period. We estimated site persistence (an occupied site stays occupied), and colonization through two subprocesses: first colonization (site creation at the time scale of the study) and recolonization (a site is colonized again after being deserted). Our model explicitly incorporated local and neighboring breeding success and conspecific density in the neighborhood. Our results provided evidence that "successful" sites have a higher persistence probability than "unsuccessful" ones. Analyses of site fidelity in marked birds and of survival probability showed that high site persistence predominantly reflects site fidelity, not immediate colonization by new owners after emigration or death of previous owners. There is a negative quadratic relationship between local density and persistence probability. First colonization probability decreases with density, whereas recolonization probability doesn't. This highlights the importance of distinguishing initial and recolonization to understand site occupancy. All dynamics varied positively with neighboring breeding success. We found evidence of a positive interaction between local and neighboring breeding success.

Keywords: *Rissa tridactyla*, habitat selection, first and re-colonization probability, site occupancy, density, breeding success

.

INTRODUCTION

Understanding the distribution of animal populations has long been central to population ecology. The vast majority of animal species live in a heterogeneous and patchy environment (e.g. at the physicochemical or the landscape level, or because food availability varies over space). Patches, or even breeding sites within patches, exhibit different characteristics, and individuals of mobile species moving actively are assumed to be able to choose the patch (or site) they occupy, unless competition prevents them from doing so (Kokko et al. 2004). As emphasized by Andreassen et al. (2002), the link between animal movement and spatial patterns of population dynamics is poorly understood. In this study, we addressed breeding site occupancy dynamics in a long-lived colonial seabird species using spatio-temporal site occupancy models (MacKenzie et al. 2006, Royle and Dorazio 2008, Bled et al. *in press*) explicitly incorporating hypotheses from behavioral ecology.

Evolutionary ecology provides a conceptual framework to address the relationship between individual decisions and habitat features (Fretwell and Lucas 1970). Breeding habitat characteristics may be associated with variation in the probability of producing offspring (a component of fitness), and ultimately fitness (Oro 2008). Different tactics of habitat selection may result in variable fitness; which is one of the premises of natural selection (phenotypic variation for a trait and a fitness function; Fairbain and Reeve 2001). In the framework of evolutionary ecology, it is assumed that individuals choose their breeding habitat based on the expected fitness in this habitat (Fretwell and Lucas 1970), but that constraints may prevent them from settling in the most favorable habitat.

Individuals that bred at least once have two options: remain at a breeding site previously occupied or change sites (i.e. bird's nest). Natural selection should favor individual tactics enhancing fitness. Assuming that individuals are free to settle where they want, individuals with breeding experience are expected to leave a breeding site they previously used if there is another site where their expected fitness is higher (even if changing sites is costly). This

hypothesis relies on three prerequisites. First, that habitat quality varies over space; if there is no spatial variability in habitat quality, 'habitat selection' is not relevant. Second, if breeding sites are chosen on the basis of achieved past fitness (related to habitat quality), reproductive experience (or any physical or social characteristic associated with it) should be repeatable over time in a given location (temporal autocorrelation; Boulinier and Lemel 1996). This condition must be met for any quality criteria observed in the past to reflect expected future fitness. Third, individuals can assess the potential fitness of breeding sites before settling, for example by prospecting several breeding sites (Danchin et al. 1991, Petit and Petit 1996). Site-specific expected fitness may be expressed through environmental, physical or social cues. The byproduct of these site characteristics and individual characteristics is realized breeding success on an occupied site. Breeding success integrates all the parameters that may determine the quality of a specific site, even parameters not directly related to the site itself (such as presence of predators in the area or food availability) (Burger 1982, Boulinier and Lemel 1996). Hereafter, site quality will refer to the probability that an individual breeds successfully on the site (i.e. high quality habitat corresponds to habitats where individuals can expect high fitness).

Several hypotheses have been proposed concerning criteria used by individuals to assess expected fitness and make choices. Individuals may directly assess site quality based on environmental and social parameters (Valone and Templeton 2002) or using their own realized fitness if they have reproduced on this site (Boulinier et al. 2001). They may also use social criteria, i.e. information created by conspecific density ("conspecifics attraction", Stamps 1988), or by their breeding success (Switzer 1997). Conspecific presence may play a part in the habitat selection process through density either positively (Serrano and Tella 2007), or negatively. Density may provide information on habitat quality (i.e., expected fitness) even if counterexamples exist (Van Horne 1983). Negative density-dependence may operate through food availability (e.g. in seabird species where feeding and nesting areas are distinct; Ashmole

and Ashmole 1967, Ricklefs 1983, Lewis et al. 2001) or nesting site availability (Kokko et al. 2004).

In the kittiwake (*Rissa tridactyla*), a cliff nesting seabird, previous work conducted in Brittany (France) has provided evidence supporting the hypothesis that individuals use their own breeding success and the success of conspecifics, or any criterion consistently associated with success but unobservable by investigators, for habitat selection at the cliff and colony scales (Danchin and Monnat 1992, Danchin et al. 1998). However, if individual movement among colonies or cliffs and the dynamics of these sub-populations are relatively well understood in kittiwakes, no work has addressed why, on average, 30% of individuals breeding on a cliff switch site within cliffs every year (Cam 1997). The hypothesis put forward to account for breeding dispersal within small spatial units is that heterogeneity in site quality within cliffs plays the same part as heterogeneity in habitat quality at a coarser spatial grain (e.g. colony) in the habitat selection process; movement among sites within cliffs may reflect individual habitat selection tactics based on variation in expected fitness. The process described at a coarser spatial grain may also hold in smaller spatial units within colonies (Cam 1997), as suggested in a ground nesting seagull species (Parejo et al. 2006). Whether ecological and evolutionary processes identified at a given spatial scale hold at another has been recognized as an important question (Wiens 1989, Oro 2008). Moreover, whether density interacts with other habitat quality criteria, including quality assessed through breeding success has been addressed in only a handful of studies (Citta and Lindberg 2007). In a colonial species breeding in dense colonies, high density may be associated only with a marginal negative effect on breeding success probability (such as, increased occurrence of fights among parents and chick loss). However, at lower density the importance of breeding success in individual decisions may vary according to the number of neighbors in the vicinity.

We developed a model of breeding site occupancy expending the approaches introduced by MacKenzie et al. (2006), Royle and Kéry (2007) and Bled et al. (*in press*). This model

encompasses time and space, and accounts for the presence of conspecifics, breeding success on the site itself and on neighboring sites. Hypotheses about factors potentially involved in individual decisions are naturally expressed using a model in which nest use probability is parameterized as a function of spatially and temporally neighboring system states. Site persistence probability (i.e. the probability of a site being occupied in consecutive years) reflects the probability of birds selecting the same site in consecutive years or new individuals replacing others on that site immediately if previous owners died or emigrated. We expect higher persistence probability associated with success of the owners of the site and of their neighbors. Importantly, we split the colonization process (i.e. the use of a previously unused site within a given time frame) into two different subprocesses: "first" colonization (at the time scale of the study) and recolonization. These subprocesses may involve different behavioral responses to environmental and social variation. Colonial species have often been shown to nest repeatedly in locations where individuals incur density related costs, but ignore nearby unoccupied locations apparently suitable (Fletcher 2006). Recolonization probability is expected to be higher than first colonization probability up to a certain density, and this relationship might be inverted at higher density.

Stamps (2001) and Safran (2004) have shown that individuals can use conspecifics past presence to choose to settle in a specific location. However, to our knowledge, consideration of both first colonization and recolonization to better understand nesting site use dynamics has never been addressed in colonial birds. Colonization as a whole process is a thoroughly studied topic. First colonization and recolonization have been extensively studied separately (Tunnicliffe et al. 1997, Kiss and Magnin 2003), but not simultaneously. Here we considered site occupancy dynamics as the result of persistence, first colonization, and recolonization after abandonment. The model we developed is also used to address whether first colonization and recolonization probabilities vary with density in a different manner, and if this distinction is

useful. We have developed models explicitly accounting for habitat selection behavior, which has been recognized as an important need in studies of habitat selection and occupancy (Boulinier and Lemel 1996).

METHODS

Data

We used data from a study led by Jean-Yves Monnat and colleagues since 1979 (continuing at present), in six colonies of black-legged kittiwakes (*Rissa tridactyla*) located in Cap Sizun (Brittany, France, 48°5'N, 4°36' W). This program is based on i) an individual survey of kittiwakes using a Capture-Mark-Recapture approach, ii) counts of breeders, and iii) evaluation of chick production in all the colonies. Each colony consists of several nesting cliffs (Cadiou et al. 1994, Naves et al. 2006), and represents a distinct geographical location. Cliffs correspond to vertical walls with homogeneous aspect; colonies consist of several walls along the coast line separated by vertical ridges or segments of the coast line not occupied by kittiwakes. Since 1979, several hundred chicks and fewer adults have been marked every year using an individual combination of plastic and metal bands. Pairs were included in analyses if they conducted nest construction up to the stage of having a cup dug in a platform of mud and grass (Maunder and Threlfall 1972). Nonbreeders (birds skipping a breeding opportunity; Cam et al. 1998) were excluded from analyses.

Each cliff has been mapped on photographs (Adobe Illustrator, 1988 Adobe systems Inc.) and nesting sites have been located on these maps: we know their coordinates using an arbitrary coordinate system on photographs. In this paper the 'site' corresponds to the location of nests within the cliff. Once a nesting site is created, i.e. has been used at least once, it remains as an available site in subsequent years. The occupancy and breeding success history of sites is known since their creation (but only since 1979). The collection of data from sites

and individuals has been more precisely described in Cadiou and Monnat (1996), but we will highlight points particularly relevant to our purpose.

We focused on a specific cliff located in a colony colonized during the study (1985-2005). It became the largest at the end of the 1990s, and is now the most populated cliff in this study area (Cam et al. 2004). Our goal was to model the colonization process at the level of the site within this cliff. Relevant data are site occupancy states from 1985 to 2005. We have seen the establishment of this cliff (i.e., establishment of breeders) in 1985. Since its creation the cliff experienced a phase of massive colonization, desertion (which coincides with the temporary extinction of the whole colony) and finally recolonization (Cam et al. 2004). The first observation of a site being occupied corresponds to its first occupation at least since the 1960's (based on count surveys conducted in the Cap Sizun area). Three hundred and twelve nesting sites have been observed over the whole study period in this specific cliff which is now a highly dense cliff where nesting sites touch one another in some areas (fig.1). Today, this cliff holds more than 1/3 of the total population of breeders in the Cap Sizun area. In addition to these nest-site occupancy data, we also know the reproductive success of each site; we used it to assess the relationship between site-specific and neighboring breeding success and site fate. Reproductive success has been encoded as 1 if individuals succeeded to produce at least one fledgling, 0 otherwise. Our sample included data from 2815 reproduction events. Since we worked with site-related information, these reproductive events concerned both marked and unmarked individuals. The mean duration during which a previously occupied nesting site is left unoccupied is 2.97 years ±2.21(±s.d.).

Statistical analyses: spatial neighborhood

To address the relationship between the state of neighboring sites (occupied, not occupied, success, failure) and site occupancy dynamics, the spatial structure of the nest "population" is required. We defined this structure based on spatial proximity of nesting sites to one another.

To assess local density and the proportion of neighboring occupied sites whose owners bred successfully, we used a connection network among sites. The network definition was based on a threshold distance over which sites are not expected to influence one another. All sites falling under this threshold distance to a nest are neighbors of this nest. This threshold was determined by the mean range of influence of a site calculated using classical covariograms on the variable "site success probability" (Bled 2006). An illustration of this area of influence is given on a map of the cliff in figure 2. Every site *j* that falls under this influence range of site *i* is considered as connected (i.e. w_{ij}=1), and conversely w_{ij}=0 if the distance between sites *i* and *j* is over this range. Based on our data and the threshold distance range, the maximum number of neighbors a site can have -based on our data and this threshold distance range- is 44, so the maximal value of density of neighboring sites of site *i* in year *t* ($D_{i,t}$) is 44. However, here we expressed this density as a percentage of maximum possible local density (i.e. $D_{i,t}$=44 corresponds to 100%). The prerequisite of spatial and temporal correlation of site success probability in the studied cliff over the study period has also been checked (Bled, 2006).

THE MODEL

Occupancy state model

We consider occupancy data obtained by repeating sampling of *i*=1,2,…,M spatial units (i.e. nesting sites), over *t*=1,2,…, T periods of time. The dynamics of the occupancy status will be accounted for by 2 parameters: persistence probability 'ϕ' (or its complement, site-specific extinction: (1-ϕ)), and colonization probability 'γ'. Let the occupancy state $Z_{i,t}$ of site *i* (i.e. nest) in year *t*. If site *i* is occupied at time *t*, then $z_{i,t}$=1, else, $z_{i,t}$=0. We were interested in the probability $\mu_{i,t}=Pr(Z_{i,t}=1|z_{i,t-1})$ (the probability that a nest site is occupied conditional on its state in the previous year). Therefore, let's define $\phi_{t,}$ the probability that a site "survives" from time *t* to *t+1*, that is, given that it was occupied at time *t*, the probability that it is occupied again at

time $t+1$, i.e. ϕ_t=Pr($Z_{i,t+1}$=1 | $z_{i,t}$=1). Where MacKenzie et al. (2006) used site-specific extinction probability (1- ϕ_t), we prefer to use its complement, i.e. persistence probability. In addition, let γ_t stand for site-specific colonization probability, from time *t* to *t+1*, i.e. γ_t =Pr($Z_{i,t+1}$=1 | $z_{i,t}$=0). Considering both site-specific persistence and colonization parameters, we can express $Z_{i,t}$ as the outcome of a Bernoulli random variable with expected value $\mu_{i,t}$

$$Z_{i,t} \mid z_{i,t-1} \sim Bern(\mu_{i,t}) \quad (1.1)$$

where

$$\begin{aligned} \mu_{i,t} &= \Pr(Z_{i,t} = 1 \mid Z_{i,t-1} = z_{i,t-1}) \\ &= z_{i,t-1}\phi_{t-1} + (1 - z_{i,t-1})\gamma_{t-1} \end{aligned} \quad (1.2)$$

Thus, the expected value of occupancy state at time *t* depends on the previous state.

This model can easily be extended to other situations. Since detection probability of reproducing individuals is close to 1.0 in kittiwake colonies in Cap Sizun (Cam et al. 1998), we do not consider any observation process incorporating detection probability.

Recolonization reparametrization

In the model described above, all the sites not occupied have the same probability of being colonized. We distinguished between the ("sub-") processes of "first colonization" (i.e. the site has never been occupied before at the time scale of the study; this is the creation of the site in the dataset), and "recolonization" (i.e. colonization of a site after a previous “extinction”). This distinction may be useful to deal with a wide range of topics where these two processes are influenced by different factors. To account for the two sub-processes of colonization, we split the colonization parameter γ_t into a first colonization parameter denoted by "γ_t", and a recolonization parameter "θ_t". We defined a new auto-covariate $A_{i,t}$ that indicates the availability of a site for first colonization: $A_{i,t}$=1 if the site has never been occupied/colonized

before (i.e. for the period T, $A_{i,t}$=1 if $\sum_{T=1}^{t} z_{i,T} = 0$), $A_{i,t}$=0 otherwise. We can express $A_{i,t}$ as the indicator function $A_{i,t} = \prod_{k=1}^{t-1}(1-z_{i,k})$, and as a consequence, sites have different colonization probabilities depending on whether $A_{i,t}$=1 or $A_{i,t}$=0, respectively, γ_t and θ_t.

The model becomes:

$$Z_{i,t+1} \mid z_{i,t} \sim Bern(\mu_{i,t+1}) \tag{1.3}$$

with

$$\mu_{i,t+1} = \phi_t z_{i,t} + \gamma_t (1-z_{i,t}) A_{i,t} + \theta_t (1-z_{i,t})(1-A_{i,t}) \tag{1.4}$$

In a site history such as h=001011, there is no first colonization in year 1 and 2. A first colonization event occurs in year 3, followed by an extinction event. In year 5 the site is recolonized and persists from year 5 to 6. In terms of probabilities, the probability of this state-history is Pr(Z=h)= (1-γ)*γ*(1-θ)*θ*ϕ.

Spatial structure

In the model as it stands, the occupancy status of each site is independent of its environment (i.e., the location of occupied sites and the spatial structure of occupancy). However, in a large number of cases, we want to consider the influence of the spatial structure of occupancy of several sites (i.e., in a given area) on the fate of individual sites. Autocorrelation in time-space distribution models can arise from two assumptions: *i*) the conditions defining the site are autocorrelated ("*exogenous spatial dependence*"); or *ii*) sites are connected through dispersal or other behaviors that lead to spatial patterning such as aggregation or regular spacing ("*endogenous spatial dependence*") (Lichstein et al. 2002). Here, we expressed the possible influence of the spatial structure on the dynamics parameters of persistence, first colonization and recolonization through the relationship between these parameters and a covariate $D_{i,t}$.

Let N_i represent the set of sites that are neighbors of the site i, and let n_i be the number of neighbors of site i (i.e. cardinality of N_i). We defined the spatio-temporal autocovariate $D_{i,t}$ as:

$$D_{i,t} = \sum_{j \in N_i} z_{j,t}.w_{ij} \quad (1.5)$$

where w_{ij} is a weight that can be used to specify a spatial relationship between a neighbor site j and the site i. Here it is based on a connection net, where w_{ij}=1 if site j is connected to site i (for example, if the distance between i and j is under the defined/determined threshold "influence" distance), and w_{ij}=0 otherwise (i.e. if site i is not in the influence area of j). The underlying assumption is that habitat quality varies over space, and that sites closer to one another or more alike than sites father apart ("exogenous spatial dependence"), and this assumption has been verified (Bled 2006). It is important to note that here $D_{i,t}$ represents local density in the vicinity of site i at time t.

We allowed the dynamics parameters to depend on this autocovariate. Consider the example of the persistence parameter ϕ. It is expressed as a function of $D_{i,t}$ as follow:

$$\text{logit}(\phi_{i,t}) = a_t + \phi_1 D_{i,t} + \phi_2 D_{i,t}^2 \quad (1.6)$$

It should be noted that ϕ is now indexed by both time period t and site i. Note that Eq. 1.6 is a quadratic function of $D_{i,t}$; this function allows ϕ to vary in a way that may lead to a peak at intermediate values of $D_{i,t}$. If the persistence parameter is a linear function of density, then the estimate of ϕ_2 should be close to zero. Moreover, in this case, the information criterion (see below) obtained for the model with a simple linear relationship and the one with a quadratic relationship should be approximately the same. We modeled $\gamma_{i,t}$ and $\theta_{i,t}$ similarly, but other approaches can be used if investigators want one or the other of the dynamics parameter to depend on the spatial structure of occupancy.

The model can be described as follow:

$$Z_{i,t+1} \mid z_{i,t} \sim Bern(\mu_{i,t+1})$$

with

$$\mu_{i,t+1} = \phi_{i,t} z_{i,t} + \gamma_{i,t}(1 - z_{i,t}) A_{i,t} + \theta_{i,t}(1 - z_{i,t})(1 - A_{i,t})$$

where

$$\begin{cases} \text{logit}(\phi_{i,t}) = a_t + \phi_1 D_{i,t} + \phi_2 D_{i,t}^2 \\ \text{logit}(\gamma_{i,t}) = b_t + \gamma_1 D_{i,t} + \gamma_2 D_{i,t}^2 \\ \text{logit}(\theta_{i,t}) = c_t + \theta_1 D_{i,t} + \theta_2 D_{i,t}^2 \end{cases}$$

depending on

$$D_{i,t} = \sum_{j \in N_i} z_{j,t} . w_{ij}$$

where a_t , b_t, and c_t are intrinsic parameters for persistence, first colonization and recolonization respectively (i.e. they reflect what happens when sites have no occupied neighbors, so that $D_{i,t} = 0$).

Habitat selection: density, site-specific and neighboring breeding success

Conspecific attraction, density — First, we accounted for the presence of nearby individuals. The underlying assumption is that this presence can influence nest site dynamics parameters via an "endogenous spatial dependence". Breeding in densely populated areas has advantages such as predator protection or information about habitat and costs related to competition for resources or increased disease transmission. When advantages overcome cost, conditions are favorable for colony creation to happen.

In our model, the quadratic relationship between dynamics parameters and local density reflects the accelerating or decelerating effect of increasing density on individual choices. It is not realistic to consider a linear relationship between persistence probability and density; this would imply a constant increase or decrease in, say, persistence probability, for a given change in density at any point of its range. For example, the Allee effect (Courchamp et al. 1999), if present, can be detected using quadratic functions. An Allee effect occurs in small populations when the decrease in density leads to a decrease of positive interactions among individuals. In

this case a decrease in local density will be associated with a decrease in dynamics parameters. At high density, a decrease in dynamics parameters decrease would reflect a phenomenon of overpopulation. Again, a quadratic model is necessary to detect such a phenomenon.

Site-specific breeding success —Reproductive success $S_{i,t}$ of an individual on site *i* (i.e. site-specific breeding success) equals to 1 if the individual produced one or more fledgings, and 0 otherwise. Our hypothesis is that breeding success should be associated with increased persistence, and failure should lead to a decrease in this parameter.

We redefined persistence probability with two intrinsic parameters a_t and a'_t :

$$\text{logit}(\phi_{i,t}) = a_t(1-S_{i,t}) + a'_t S_{i,t} + \phi_1 D_{i,t} + \phi_2 D_{i,t}^2$$

a'_t is the persistence probability of a "successful" site (i.e., the owners of the site bred successfully), and a_t the persistence probability of a "failed" site. If our hypothesis is supported, the estimate of a'_t should be higher than a_t .

Neighboring breeding success — Breeding success of neighbors or individuals in an area has been suggested to reflect the quality of the area, and therefore of a site in this area (Danchin et al. 1998). Under this hypothesis, the success in the vicinity of a specific site is expected to influence all dynamics parameters. The underlying assumption is that the surrounding neighborhood of breeders and their success influences nest site selection, retention and creation ("endogenous spatial dependence"). If we consider the previous definition of persistence, we decided to use a parameter ϕ_S that will weight the ratio of neighboring sites that are successful over the total number of occupied neighbor sites. Such an approach has been used in this population at the spatial scale of the cliff (Aubry et al. 2009). This ratio $\tau_{i,t}$ corresponds to the relative success of the neighborhood and is equal to $\frac{N_{S_{i,t}}}{N_{S_{i,t}} + N_{F_{i,t}}}$ where $N_{S_{i,t}}$ is the number of successful neighbor sites, and $N_{F_{i,t}}$ the number of unsuccessful neighbors. Their sum is the

total number of occupied sites neighboring site *i* at time *t*. Then, the model is extended to include this variable as follows:

$$\text{logit}(\phi_{i,t}) = a_t(1-S_{i,t}) + a'_t S_{i,t} + \phi_1 D_{i,t} + \phi_2 D_{i,t}{}^2 + \phi_S \tau_{i,t}$$

Our hypothesis is that persistence probability should increase with the proportion of successful sites, i.e. ϕ_S should be positive.

In addition, we expressed first colonization $\gamma_{i,t}$ and recolonization $\theta_{i,t}$ as:

$$\text{logit}(\gamma_{i,t}) = b_t + \gamma_1 D_{i,t} + \gamma_2 D_{i,t}{}^2 + \gamma_S \tau_{i,t}$$

and

$$\text{logit}(\theta_{i,t}) = c_t + \theta_1 D_{i,t} + \theta_2 D_{i,t}{}^2 + \theta_S \tau_{i,t}$$

Interactions among site-specific success, neighboring success rate and density — We considered the hypothesis that the relationship between site-specific breeding success and site occupancy might vary according to success of neighbors. Namely, site-specific breeding success may no longer be taken into account by the individual occupying a site once a given level of breeding success of neighbors has been reached (poor productivity of neighbors), whereas it may be taken into account in different productivity contexts (intermediate or high proportion of successful neighbors). In other words, one criterion used to assess habitat quality may be ignored under some circumstances (van Bergen et al. 2004). Such a hypothesis can be addressed using an interaction between site-specific and neighboring breeding success on persistence probability. Moreover, recent theory suggests that population density may influence how individuals use social information; this alters the potential costs and benefits of making decisions based on social cues (Fletcher 2007). These components of behavioral responses to social conditions might be reflected in the relationship between site occupancy and the social context. We designed our model to include interactions between neighboring breeding success and density (respectively ϕ_{int1}, γ_{int} and θ_{int} for persistence, first colonization

and recolonization parameters), and between site-specific and neighboring breeding success ($\phi_{\text{int}2}$).

$$\text{logit}(\phi_{i,t}) = a_t(1-S_{i,t}) + a'_t S_{i,t} + \phi_1 D_{i,t} + \phi_2 D_{i,t}{}^2 + \phi_S \tau_{i,t} + \phi_{\text{int}1} D_{i,t}.\tau_{i,t} + \phi_{\text{int}2} S_{i,t}.\tau_{i,t}$$
$$\text{logit}(\gamma_{i,t}) = b_t + \gamma_1 D_{i,t} + \gamma_2 D_{i,t}{}^2 + \gamma_S \tau_{i,t} + \gamma_{\text{int}} D_{i,t}.\tau_{i,t}$$
$$\text{logit}(\theta_{i,t}) = c_t + \theta_1 D_{i,t} + \theta_2 D_{i,t}{}^2 + \theta_S \tau_{i,t} + \theta_{\text{int}} D_{i,t}.\tau_{i,t}$$

A list of all parameters used and estimated in the model is provided in appendix 1.

SURVIVAL AND SITE FIDELITY IN MARKED INDIVIDUALS

Site persistence probability reflects the probability of birds selecting the same site in consecutive years or new individuals replacing others on the site immediately if previous owners died or emigrated. We analyzed site fidelity (i.e. the owner of a site breeds again on the same site in the following year) and survival probability in marked birds to assess whether high site persistence probability primarily reflects fidelity of previous owners or immediate colonization by new owners after death of previous owners or emigration.

To assess the contribution of site fidelity by previous owners and colonization by new owners to site persistence probability, we addressed the relationship between the probability of re-nesting on the same site, or of surviving; and individual breeding success, the success of neighbors, and density. We expect higher persistence probability associated with success of the owners of the site and of their neighbors, either through site fidelity, or immediate attraction of new breeders. To address whether low persistence predominantly reflects emigration or death of owners, survival probability was used to measure the contribution of mortality and emigration to low site persistence probability.

Comparison of site fidelity, survival, and site persistence probability estimates made in similar conditions (i.e., site-specific and neighboring success, density) should provide insight into whether high site persistence predominantly reflects high site fidelity or high colonization

by new owners, and whether low persistence probability predominantly reflects death of owners or their decision to leave.

Previous studies have provided evidence that recapture probability is virtually equal to 1 after recruitment in this population (Cam et al. 2005). Consequently, both site fidelity and survival probability can be estimated using Generalized Linear Models (GLMs) without accounting for detectability. We used GLMs to address the relationship between year, sex, site-specific success, neighboring success, local density and site fidelity on the one hand, and with survival on the other hand. Sex was considered because males are territorial in this species, and females whose own mate died rarely manage to durably evict intruding males, potentially leading site fidelity to vary depending on sex.

BAYESIAN ANALYSIS

Models containing a large number of latent binary variables have previously been fitted using Bayesian methods implemented in the software program WinBUGS (Wintle et al. 2005, Royle and Kéry, 2007, Bled et al. *in press*). Models were implemented in the free software package WinBUGS 1.4 (Lunn et al. 2000) called from R 2.6.2 (R Development Core Team 2008) using the R library R2WinBUGS (Sturtz et al. 2005). We ran 3 chains using non informative priors and based our inference on 100 000 samples from the posterior distribution of parameters, after 20 000 discarded iterations. The WinBUGS code for the selected model is provided in Appendix 2.

Model selection was conducted using the Deviance Information Criterion (DIC, Spiegelhalter et al. 2002) given by WinBUGS. For analyses of site fidelity and survival probabilities, model selection was conducted using a downward selection procedure based on Akaike's Information Criterion (AIC, Tables 2 and 3, respectively).

RESULTS

The model with the largest support (model probability of 0.56, Table 1) includes site-specific breeding success, a quadratic relationship between density and site persistence $\phi_{(i,t)}$, a linear relationship between first colonization $\gamma_{(i,t)}$ and density, but does not include any relationship between density and recolonization $\theta_{(i,t)}$. In the selected model, there is a relationship between neighboring success and all dynamic parameters (site persistence, first colonization and recolonization). There is an interaction between site-specific and neighboring breeding success but no interaction between density and neighboring success. The parameterization of this model is designed to split the colonization process into first colonization and recolonization. The next closest model, with a model probability equal to 0.37, also includes an interaction between density and neighboring breeding success only for the persistence parameter (Table 1).

Relationship between site-specific breeding success and persistence

Site persistence probability is higher when reproduction on a site is successful than when the owners fail. The mean probability of persistence of a successful site over the study period equals 0.95 (±0.01) (± s.d.) while the mean persistence probability of an unsuccessful site is equal to 0.82 (±0.02). These estimates correspond to mean values of density and success rate in the neighborhood during the study. If the owner of a site bred successfully, the site is more likely to be reused in the following year than to be vacant, either because the owners return and breed on the same site, or because new individuals occupy the site.

Relationship between conspecifics presence (through density) and dynamics parameters

Figure 3 shows mean estimated intrinsic parameters over the period of study as a function of local density around the site *i*. The relationship between persistence probability and density

is negative quadratic. Persistence probability is low at low density, higher at intermediate density, and decreases at high values of density (estimates of ϕ_2 and ϕ_1 equal to -0. 37 (±0.08) and 1.49 (±0.30), respectively).

A linear negative relationship was found between the 'first colonization' parameter $\gamma_{(i,t)}$ and density. The probability of creation of "new" sites (at the scale of the study period) slightly increases (γ_1 = 0.11 ±0.1) with density.

Recolonization does not vary as a function of density. For an average neighboring success rate, a site has approximately 33% chances of being recolonized. Over the whole density range, recolonization probability stays higher than first colonization probability indicating that (for an average success rate) site re-use is more frequent than settlement on a previously unoccupied site.

Relationship between neighboring breeding success and dynamics parameters

Figure 4 corresponds to mean values of estimated intrinsic parameters over the study period, as a function of mean values of the success rate in the neighborhood of site *i*. All the dynamics parameters (persistence, first colonization, recolonization probabilities) increase with the success rate of the neighborhood. Slope coefficients, on the logit scale, corresponding to neighbor success rate are: 1.03 (±0.41) for the persistence parameters, 0.59 (±0.21) for the first colonization parameter and 0.08 (±0.28) for the recolonization parameter. Persistence varies positively with neighboring breeding success, site creation (through first colonization parameter) is higher where neighbor breeding success rate is high, while this relationship is less marked for the reuse of previously occupied site. Again, site reuse is higher over the whole range of neighboring success rate than site creation.

Interactions among covariates

We found evidence of an interaction between neighboring breeding success and site-specific breeding success. The interaction is positive ($\phi_{int\,2}$ =2.14 ± 0.55), indicating that when neighboring sites are successful (i.e., the owners breed successfully), the site *i* itself has a higher probability of being successful, which might simply reflect that higher-quality sites form clusters in local cliff areas.

Owner's fidelity to site

In the selected model (with an AIC weight equal to 0.64, table 2), the probability of site fidelity by marked individuals varies positively with the owner's success on this site (estimated slope parameter = 3.48 ±0.72), the neighboring success rate (estimated slope parameter = 1.29 ±1.29) and density (estimated slope parameter = 0.25 ±0.15). There is also a year effect and a relationship between fidelity and sex, with male having a higher fidelity than females (estimated slope parameters = 0.44 ±0.16). Interactions between owner's success and both year and density (estimated slope parameters = -0.42 ±0.21) were selected.

While the estimated site fidelity probability in conditions of high site-specific (S=1) and neighboring success (τ= 95 %), and intermediate density (D= 50 %) is equal to 0.9, in the same conditions, site persistence probability is equal to 0.99. In conditions of low site-specific (S=0) and neighboring success (τ= 10 %) and intermediate density (D= 50 %), estimated fidelity and persistence probabilities are equal to 0.86 and 0.81, respectively. Estimates of site persistence and fidelity probabilities are very close, and vary in the same way when conditions change. These results indicate that the main process underlying site persistence is fidelity, not immediate recolonization after abandonment. In addition, fidelity increases with owner's success, neighboring success and density.

Owner's survival

In the best model (AIC weight = 0.47, table 3), there is a year effect and a relationship between survival and sex of marked individuals, males exhibiting a lower survival than females (estimated slope parameter = -0.24 ±0.14). Survival probability varies positively with site-specific success (estimated slope parameter = 0.27 ±0.15).

We did not find evidence that survival probability varies with neighboring success or density, indicating that variation of persistence probabilities with these habitat features does not result from variation in survival probability. When site-specific (S=0) and neighboring success (τ= 10 %) are low, and density is intermediate, (D= 50 %), estimated survival and persistence probabilities are equal to 0.83 and 0.81, respectively. In conditions of high site-specific breeding success (S=1), high neighboring success (τ= 95 %) and intermediate density (D= 50 %), estimated survival and persistence probabilities are 0.86 and 0.99, respectively. Because there is no covariation of neighboring success and survival probability on the one hand, and density and survival on the other hand, and because survival probability is relatively high whether the individual experiences breeding success or failure (while persistence probability varies more substantially), we conclude that site abandonment is not mainly due to death of the owner.

DISCUSSION

Relationship between site-specific success and site persistence

Persistence probability of a site is higher when individuals occupying the site were successful. Here, the probability of being occupied in the following year is almost 17% higher for a successful site compared to unsuccessful ones. Assuming that site fidelity plays a substantial part in nesting site persistence, as corroborated by analyses of owner's site fidelity, this is consistent with a hypothesis developed in the habitat selection framework: individual breeding success influences the choice of an individual to stay or not (Boulinier et al. 2001).

This is also consistent with the hypothesis that the intrinsic characteristics of the site itself associated with breeding success influence individual decisions (Parejo et al. 2006). Previous studies have shown that site fidelity is higher on productive sites (Cam 1997) and cliffs (Danchin et al. 1998). In addition, mean estimated survival probability is close to 0.80 in breeders (Cam et al. 2005), and higher in successful breeders than in unsuccessful breeders (Cam et al. 1998). Our analyses complement these results: high site persistence probability corresponds to high probability of site fidelity in similar conditions of site-specific success, neighboring success, and density, and to high survival probability. Consequently, desertion or death of previous owners followed by immediate recolonization is not the main explanation for high site persistence probability. In addition, low persistence probability does not predominantly reflect mortality, but emigration of the owners.

Relationship between conspecifics presence through density and dynamics parameters

The site persistence relationship with density is characterized by a negative quadratic shape. What is observed at low density can be compared to an Allee effect (Courchamp et al. 1999). The decrease in positive interactions at low densities can be associated with decreased social attraction (Stephens et al. 1999). Conversely, at high density, we observe a phenomenon similar to overpopulation. In such dense colonies, nests are sometimes so close to one another that there is contact and aggressive behavior between incubating adults occur. Also, when the number of nests in an area increases, the probability of settling on a "poor-quality" site may increase. Site persistence at intermediate density may result from trade-off between these two phenomena. Conspecifics presence can be used as a cue to assess patch quality (Serrano and Tella 2003), and its use as a criterion taken into account in the choice of breeding habitat has already been shown in several species at scales such as the colony or even the cliff (Muller et al. 1997, Danchin et al. 1998). In this study, we provided evidence that it can also be used at nesting site scale inside habitat patches where breeding success varies. Moreover, it should be

noted that the estimated persistence probability is higher than 0.5 over the whole density range, which means that a site is more likely to remain occupied.

Comparison of estimates of "first colonization" and recolonization provided evidence that recolonization probability is higher than site creation. This means that some sites tend to be occupied first, and suggests a preferential use of some sites. Moreover, first colonization probability varies as a function of density while recolonization does not. Higher-quality sites are expected to be chosen first (Newton and Marquiss 1982), and when they get deserted for any reason by individuals (death or dispersal), abandoned sites are expected to be reoccupied first, before new breeding sites are "created". Fairweather and Coulson (1995) showed that even if nesting sites have been made unusable because of modifications of the environment, individuals kept trying to breed on the sites they previously used. When sites became accessible again a couple of years later, individuals returned to breed there. In our study, the fact that first colonization probability increases with density might indicate that, while some sites are occupied preferentially, when these sites are not accessible, individuals settle in denser area, which can be explained by conspecifics attraction. It remains to be investigated whether the differences in the variations in first versus recolonization as density changes are due to different types of individuals making different decisions: other studies suggest such cases for young vs. old individuals, first time breeders vs. experienced ones, immigrants vs. emigrants (Aparicio et al. 2007, Krištín et al. 2007, Arlt and Pärt 2008, Parejo et al. 2008). Splitting the global process of colonization into two sub-processes (first and re-colonization) is important to understand how "colonization" works, since they do not vary in the same way with local conditions (i.e. density).

Relationship between neighboring breeding success and dynamics parameters

First colonization and recolonization probabilities increase with the proportion of successful neighboring sites. This is consistent with previous results obtained at larger spatial

scales (Danchin et al. 1998, Doligez et al. 2003). However, our results also show that site occupancy patterns and dynamics vary as a function of site-specific breeding success and density. This might reflect the use of sources of information on habitat quality such as personal and public information and conspecific presence-related information.

Individuals have to consider neighboring breeding success information in light of other indications provided by personal information, such as their own success, and conspecifics presence. When an individual breeds successfully in an area where most neighbors failed, it is more likely to stay and reproduce again on this site than unsuccessful individuals in low productivity areas (Danchin et al. 1998, this study). This would translate into higher persistence probability for successful sites in higher-quality areas. In the study population, massive failure (i.e., situations where entire areas exhibit low productivity) is mostly due to external factors such as predation on eggs by corvids (Cam et al. 2004). In contrast, when an individual fails in a relatively productive area, failure is likely to result from individual characteristics (or to issues with the pair), not from external causes. When an individual is unsuccessful in an area with high failure rate, the "cause" of failure is unclear: individual characteristics, characteristics of the pair, or biotic environmental factors? In terms of future fitness, coming back to breed on the same site may lead to high probability of failure if external factors were responsible of loss of the clutch (in environment whose quality is temporally auto-correlated). Personal and public information convey two different elements for the choice of a breeding site. Site occupancy, as the result of individuals' decisions, reflects the potential use of all of these three sources of information (personal and public information and conspecific presence).

Interactions among covariates

Our selected model included the interaction between site-specific and neighboring breeding success on nest persistence probability ($\phi_{\text{int}2} = -2.11\ (\pm 0.55)$). Danchin et al. (1998) have shown that there is an interaction between individual success and the proportion of successful neighbors in models of dispersal probability among cliffs within the study area. They concluded that at this spatial scale "individual breeding performance did not appear to influence the probability of dispersing for birds breeding on cliffs with high local reproductive success, whereas individual breeding performance has a strong effect on dispersal for birds that bred on cliffs with lower local reproductive success". Boulinier et al. (2008) showed, with experimental manipulation (eggs removal) in a Norwegian kittiwake colony, that there is an interaction between personal and conspecific success. They also showed that individuals that lost their eggs in successful areas would attend their nest and have a higher probability of coming back the following year than individuals that lost their eggs in areas where their neighbors also failed. Here we also found evidence of an interaction between site-specific and neighboring breeding success on site persistence probability. Both failed and successful sites have a high persistence probability in a successful neighborhood, but the difference in persistence probability between the two types of sites increases as the social context deteriorates (the proportion of successful sites decreases in the neighborhood) (fig. 4). In low-productivity contexts, persistence probability varies substantially with the success of the owner of the site (failed and successful sites show unambiguously different persistence probability), whereas the magnitude of the difference between persistence probability of successful and failed sites is small in productive areas

Model development and potential applications

The model used here is highly flexible and can be modified to address other, more general questions. First, here, no observation process has been included in the model, however, applications to study systems with imperfect detection probability can be handled by extending the hierarchical model (Royle and Kéry 2007, Bled et al. *in press*). Second, the main features of the model that can be modified are (i) underlying functions for dynamics parameters, (ii) weights used to establish the spatial structure, (iii) and the fact that parameters are static or time/space varying.

Despite the obvious implications of time-space processes for both basic and applied ecology, the statistical framework of modeling of spatio-temporal occupancy systems is not well developed yet, even if some recent efforts have been made (Zhu et al. 2005; Hooten and Wikle 2008) or Zhu et al. (2005). Here, it is possible to modify the model both on the time and the spatial scales by allowing static parameter to vary over time or space. Assumptions about time or space stationarity can be easily relaxed. It should be noted that the underlying assumption that there is no relationship between the cumulative number of years unoccupied and the probability that a site is recolonized can also be relaxed (i.e., long-term memory processes). Last, we can set w_{ij} in order to weight $D_{i,t}$ by the inverse distance of sites j to i, or any other way depending on the decisions made to define the connections network, and that seems more relevant according to the studied topic.

Finally, we have provided evidence of a difference in the relationships between population conditions and first colonization on the one hand, and recolonization dynamics on the other hand . As previously emphasized, this distinction has rarely been made previously (Bled et al., *in press*). The approach used to model these processes can be applied to other questions where this separation is relevant, such as the case of invasive species, or the study of systems that

undergo events of recolonization after major ecological events (sudden pollution, wildfire or flood).

Our work integrates different hypotheses developed in behavioral ecology to account for individual choices and explain local population dynamics through a site occupancy approach. Site occupancy approaches provide demographic parameter estimates for predictive models incorporating habitat characteristics, and allows assessment of common assumptions made in metapopulation models, e.g. whether a local population is a source or a sink (Hanski and Gilpin 1997). Accounting for individual behavioral decisions in site occupancy studies may enhance our ability to develop models of population and metapopulation dynamics explicitly incorporating ecological and evolutionary processes, as opposed to phenomenological models. It would be possible to compare the estimates of dynamics parameters among different species and/or scales. Importantly, site occupancy models can be modified to incorporate other classes of variables, such as the distance to the nearest neighbor, age, experience or social rank of site owners, and address other hypotheses about habitat selection in different species.

AKNOWLEDGEMENTS

The authors thank Marc Kéry, Beni Schmidt, Robert Dorazio, Jérôme Dupuis, Beth Gardner, Adam Green, Lise Aubry, David Koons, Sergine Ponsard, Michael Runge, Roger Pradel and Peter Pearman for discussions and comments on earlier versions of this manuscript. The authors are also grateful to an anonymous reviewer, Michael Erwin and Walter Koenig for their helpful comments. We are grateful to Jean-Yves Monnat, who has been in charge of the monitoring program of kittiwakes in the Iroise Sea for 32 years, and provided a large proportion of the data used in the manuscript. We also thank the Conseil Général du Finistère allowing us to work in the nature reserve of Goulien Cap Sizun (Brittany, France), managed by Bretagne Vivante - Société pour l'Étude et la Protection de la Nature en Bretagne. We express

our deep gratitude to all the people who contributed to data collection over almost three decades, and especially to Jacques Nisser (Office National de la Chasse et de la Faune Sauvage) for his help with fieldwork.

LITERATURE CITED

Andreassen, H. P., N. C. Stenseth, and R. A. Ims. 2002. Dispersal behaviour and population dynamics of vertebrates. Pages 237–256 in Bullock J. M., R. E. Kenward and R. S. Hails, editors. Dispersal Ecology. Blackwell Publishing, Malden, USA.

Aparicio, J. M., R. Bonal, and A. Muñoz. 2007. Experimental test on public information use in the colonial Lesser Kestrel. Evolutionary Ecology 21:783–800.

Arlt, D., and T. Pärt. 2008. Post-breeding information gathering and breeding territory shifts in northern wheatears. Journal of Animal Ecology 77:211–219.

Ashmole, N. P., and M. J. Ashmole. 1967. Comparative feeding ecology of sea birds of a tropical oceanic island. Peabody Museum of Natural History, Yale University, Bulletin 24. Peabody Museum of Natural History, New Haven, Connecticut, USA.

Aubry, L., E. Cam, and Monnat J.-Y. 2009. Habitat selection, Age-specific recruitment and reproductive success in a long-lived seabird. Pages 365–393 in Thomson D.L., E. G. Cooch and M. J. Conroy, editors. Modeling Demographic Processes in Marked Populations. Springer, New York, New York, USA.

Bled, F. 2006. Sélection de l'habitat de reproduction au sein de la falaise chez un oiseau marin longévif: *Rissa tridactyla*. Master 2 Biodiversité Ecologie Evolution. Université Paul Sabatier, Toulouse, France.

Bled, F., J.A. Royle, and E. Cam, (*in press*). Hierarchical modeling of an invasive spread: case of the Eurasian collared-dove *Streptopelia decaocto* in the USA. Ecological Application

Boulinier, T., and J.-Y. Lemel. 1996. Spatial and temporal variations of factors affecting breeding habitat quality in colonial birds: some consequences for dispersal and habitat selection. Acta Oecologica 17:531–552.

Boulinier, T., K.D. McCoy, and G. Sorci. 2001. Dispersal and parasitism. Pages 169–179 in Clobert J., E. Danchin, A. A. Dhondt and J. D. Nichols, editors. Dispersal. Oxford University Press, Oxford, UK.

Boulinier, T., K. D. McCoy, N. G. Yoccoz, J. Gasparini, and T. Tveraa. 2008. Public information affects breeding dispersal in a colonial bird: kittiwakes cue on neighbours. Biology Letters 4:538–540.

Burger, J. 1982. The role of reproductive success in colony-site selection and abandonment in Black Skimmers (*Rynchops niger*). The Auk 99:109–115.

Cadiou, B., J.-Y. Monnat, and E. Danchin. 1994. Prospecting in the kittiwake, *Rissa tridactyla*: Different behavioural patterns and the role of squatting in recruitment. Animal Behaviour 47:847–856.

Cadiou, B., and J.-Y. Monnat. 1996. Parental attendance and squatting in the kittiwake *Rissa tridactyla* during the rearing period. Bird Study 43:164–171.

Cam, E. 1997. Reproduction intermittente et corrélations phénotypiques entre composantes de la fitness. Cas de la mouette tridactyla *Rissa tridactyla*. PhD Thesis in Ecology. Université de Paris VI. Paris, France. 156p.

Cam, E., J.E. Hines, J.-Y. Monnat, J.D. Nichols, and E. Danchin. 1998. Are adult nonbreeders prudent parents? The Kittiwake model. Ecology 79:2917–2930.

Cam, E., J.-Y. Monnat, and J.A. Royle. 2004. Dispersal and individual quality in a long lived species. Oïkos 106:386–398.

Cam, E., E. G. Cooch, and J.-Y. Monnat. 2005. Earlier recruitment or earlier death? On the assumption of homogeneous survival in recruitment studies. Ecological Monographs 75:419–434.

Citta, J.J., and M.S. Lindberg. 2007. Nest-site selection of passerines: effects of geographic scale and public and personal information. Ecology 88:2034–2046.

Courchamp, F., T. Clutton-Brock, and B. Grenfell. 1999. Inverse density dependence and the Allee effect. Trends in Ecology and Evolution 14:405–410.

Danchin, E., B. Cadiou, J.-Y. Monnat, and R. R. Estrella. 1991. Recruitment in long-lived birds : conceptual framework and behavioural mechanisms. Acta XX congressus internationalis ornithologici 1641–1656.

Danchin, E., and J.-Y. Monnat. 1992. Population dynamics of two neighboring kittiwake *Rissa tridactyla* colonies. Ardea 80:171–180.

Danchin, E., T. Boulinier, and M. Massot. 1998. Conspecific reproductive success and breeding habitat selection : implications for the study of coloniality. Ecology 79:2415–2428.

Doligez, B., C. Cadet, E. Danchin, and T. Boulinier. 2003. When to use public information for breeding habitat selection ? The role of environmental predictability and density dependence. Animal Behaviour 66:973–988.

Fairbain, D. J., and J. P. Reeve. 2001. Natural Selection. Pages 29–43 in Fox, C. W., D. A. Roff, D. J. Fairbain, editors. Evolutionary Ecology. Oxford University Press, Oxford, UK. .

Fairweather, J. A., and J. C. Coulson. 1995. The influence of forced site change on the dispersal and breeding of the black-legged kittiwake *Rissa tridactyla*. Colonial Waterbirds 18:30–40.

Fletcher, R. J. 2006. Emergent properties of conspecific attraction in fragmented landscapes. The American Naturalist 168:207–219.

Fletcher, R. J. 2007. Species interactions and population density mediate the use of social cues for habitat selection. Journal of Animal Ecology 76:598–606.

Fretwell, S. D., and H. J. Lucas. 1970. Ideal free distribution. Acta Biotheory 19:16–21.

Hanski, I. A., and M. E. Gilpin (eds). 1997. Metapopulation biology. Ecology, Genetics, and Evolution. Academic Press. San Diego, California, USA. 512p.

Hooten, M. B., and C. K. Wikle. 2008. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. Environmental and Ecological Statistics 15:59–70.

Kiss, L., and F. Magnin. 2003. The impact of fire on some Mediterranean land snail communities and patterns of post-fire recolonization. Journal of Molluscan Studies 69:43–53.

Kokko, H., M. P. Harris, and S. Wanless. 2004. Competition for breeding sites and site-dependent population regulation in a highly colonial seabird, the common guillemot *Uria aalge*. Journal of Animal Ecology 73:367–376.

Krištín, A., H. Hoi, F. Valera, and C. Hoi. 2007. Philopatry, dispersal patterns and nest-site reuse in lesser grey shrikes (*Lanius minor*). Biodiversity and Conservation 16:987 - 995.

Lewis, S., T. N. Sherratt, K. C. Hamer, and S. Wanless. 2001. Evidence of intra-specific competition for food in a pelagic seabird. Nature 412:816–819.

Lichstein, J. W., T. R. Simons, S. A. Shriner, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs 72:445–463.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10:325–337.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy Estimation and Modeling. Academic Press. San Diego, California, USA. 324p.

Maunder, J. E., and W. Threlfall. 1972. The breeding biology of the black-legged kittiwake in Newfoundland. The Auk 89:789–816.

Muller, K. L., J. A. Stamps, V. V. Krishnan, and N. H. Willits. 1997. The effects of conspecific attraction and habitat quality on habitat selection in territorial birds (*Troglodytes aedon*). The American Naturalist 150:650–661.

Naves, L. C., J.-Y. Monnat, and E. Cam. 2006**.** Breeding performance, mate fidelity, and nest site fidelity in a long-lived seabird: behaving against the current? Oikos 115:263–276.

Newton, M., and I. Marquiss. 1982. Habitat preference in male and female sparrowhawks *Accipiter nisus*. Ibis 124:324–328.

Oro, D. 2008. Living in a ghetto within a local population: an empirical example of an ideal despotic distribution. Ecology 89:838–846.

Parejo, D., D. Oro, and E. Danchin. 2006. Testing habitat copying in breeding habitat selection in a species adapted to variable environments. Ibis 148:146–154.

Parejo, D., E. Danchin, N. Silva, J. F. White, A. N. Dreiss, and J. M. Avilés. 2008. Do great tits rely on inadvertent social information from blue tits? A habitat selection experiment. Behavioral Ecology and Sociobiology 62:1569–1579.

Petit, L. J., and D. R. Petit. 1996. Factors governing habitat selection by prothonotary warblers : field tests of the Fretwell-Lucas models. Ecological Monographs 66:367–387.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ricklefs, R. E. 1983. Some considerations on the reproductive energetic of pelagic seabirds. Studies in Avian Biology 8:84–94.

Royle, J. A., and M. Kéry. 2007. A Bayesian state-space formulation of dynamic occupancy models. Ecology 88:1813–1823.

Royle, J.A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology. Academic Press, London UK, 444p.

Safran, R. J. 2004. Adaptative site selection rules and variation in group size of barn swallows: individual decisions predict population patterns. The American Naturalist 164:121–131.

Serrano, D., and J. L. Tella. 2003. Dispersal within a spatially structured population of lesser kestrel: the role of spatial isolation and conspecific attraction. Journal of Animal Ecology 72:400–410.

Serrano, D., and J. L. Tella. 2007. The role of despotism and heritability in determining settlement patterns in the colonial lesser kestrel. The American Naturalist 169:53–67.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van der Linde. 2002. Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society: Series B 64:583–640.

Stamps, J. A. 1988. Conspecific attraction and aggregation in territorial species. The American Naturalist 131:329–347.

Stamps, J. 2001. Habitat selection by dispersers: integrating proximate and ultimate approaches. Pages 230–242 in Clobert J., E. Danchin, A. A. Dhondt and J. D. Nichols, editors. Dispersal. Oxford University Press, Oxford, UK.

Stephens, P. A., W. J. Sutherland, and R. P. Freckleton. 1999. What is the Allee effect? Oikos 87:85–190.

Sturtz, S., U. Ligges, and A. Gelman. 2005. R2WinBUGS: A Package for Running WinBUGS from R. Journal of Statistical Software 12:1–16.

Switzer, P. V. 1997. Past reproductive success affects future habitat selection. Behavioral Ecological Sociobiology 40:307–312.

Tunnicliffe, V., R. W. Embley, J. F. Holden, D. A. Butterfield, G. J. Massoth, and S. Juniper. 1997. Biological colonization of new hydrothermal vents following an eruption on Juan de Fuca Ridge. Deep-Sea Research I 44:1627–1644.

Valone, T. J., and J. J. Templeton. 2002. Public information for the assessment of quality: a widespread social phenomenon. Philosophical Transactions of the **Royal Society B** 357:1549–1557.

van Bergen, Y., I. Coolen, and K. N. Laland. 2004. Nine-spined sticklebacks exploit the most reliable source when public and private information conflict. Proceedings of the Royal Society of London B 271:957–962.

van Horne, B. 1983. Density as a misleading indicator of habitat quality. The Journal of Wildlife Management 47:893–901.

Wintle, B. A., R. P. Kavanagh, M. A. McCarthy, and M. A. Burgman. 2005. Estimating and dealing with detectability in occupancy surveys for forest owls and arboreal marsupials. Journal of Wildlife Management 69:905–917.

Zhu, H., H.-C. Huang, and C.-T. Wu. 2005. Modeling spatial-temporal binary data using Markov random fields. Journal of Agricultural, Biological, and Environmental Statistics 10:212–225.

Table 1. Parameterization of the different models and Deviance Information Criterion values. *Plus sign indicates that the corresponding element is included in the model. Relationship between density and dynamics parameters: '2','1','0' indicate quadratic, linear and no relationships respectively. Bold DIC is the selected model.*

Model elements											DIC	Normalized DIC weight
First/Re colonization differentiation (γ/θ)	Site-specific breeding success ($S_{i,t}$)	Relationship between density ($D_{i,t}$) and $\varphi_{i,t}$	$\gamma_{i,t}$	$\theta_{i,t}$	Relationship between neighboring success ($\tau_{i,t}$) and $\varphi_{i,t}$	$\gamma_{i,t}$	$\theta_{i,t}$	Interaction between density and neighboring breeding success for $\varphi_{i,t}$	$\gamma_{i,t}$	$\theta_{i,t}$	Interaction between site-specific and neighboring success		
		2	2	2								4285.7	0
	+	2	2	2								4163.5	0
		2	2	2	+	+	+					4268.2	0
	+	2	2	2	+	+	+					4258.4	0
	+	2	2	2	+	+	+				+	4147.3	0
	+	2	2	2	+	+	+	+	+	+	+	4142.8	0
	+	2	2	2	+	+	+	+	+	+		4143.9	0
+		2	2	2								4082.6	0
+	+	2	2	2								3970.9	0
+		2	2	2	+	+	+					4063.3	0
+	+	2	2	2	+	+	+					3957.6	0
+	+	2	2	2	+	+	+				+	3952.4	0
+	+	2	2	2	+	+	+	+	+	+	+	3947.5	0
+	+	2	2	2	+	+	+	+	+	+		3955.1	0
+	+	1	2	2	+	+	+	+	+	+	+	3965.3	0
+	+	2	1	2	+	+	+	+	+	+	+	3963.5	0
+	+	2	2	1	+	+	+	+	+	+	+	3959.6	0
+	+	1	1	2	+	+	+	+	+	+	+	3965.9	0
+	+	1	2	1	+	+	+	+	+	+	+	3954.7	0
+	+	2	1	1	+	+	+	+	+	+	+	3944.6	0
+	+	1	1	1	+	+	+	+	+	+	+	3957.4	0
+	+	2	0	1	+	+	+	+		+	+	3944.8	0
+	+	2	1	0	+	+	+	+	+		+	3941.3	0.03
+	+	2	0	0	+	+	+	+			+	3946.6	0
+	+	2	1	0		+	+		+		+	3945.9	0
+	+	2	1	0	+		+	+			+	3952.0	0
+	+	2	1	0	+	+		+	+		+	3941.1	0.03
+	+	2	1	0			+				+	3953.4	0
+	+	2	1	0		+			+		+	3946.7	0
+	+	2	1	0	+			+			+	3961.9	0
+	+	2	1	0							+	3957.9	0
+	+	2	1	0	+	+	+		+		+	3967.2	0
+	+	2	1	0	+	+	+	+			+	**3935.9**	**0.37**
+	+	2	1	0	+	+	+				+	**3935.1**	**0.56**

Table 2. Nesting site fidelity in the black-legged Kittiwake. Selection results (based Akaike's Information Criteria) for models including site-specific breeding success *S*, local density *D*, neighboring breeding success rate τ, year *t* and sex as covariates. *Bold AIC is the selected model.*

Model	AIC	AIC weight
Fidelity ~ S + t + S*t + sex + τ + D + S*D + S* τ + sex*S + sex*t	1252.9	0
Fidelity ~ S + t + S*t + sex + τ + D + S*D + S* τ + sex*S	1230.1	0.12
Fidelity ~ S + t + S*t + sex + τ + D + S*D + S* τ	1228.6	0.24
Fidelity ~ S + t + S*t + sex + τ + D + S*D	**1226.7**	**0.64**

Table 3. Survival in the black-legged Kittiwake. Selection results (based Akaike's Information Criteria) for models including site-specific breeding success *S*, local density *D*, neighboring breeding success rate τ, year *t* and sex as covariates. *Bold AIC is the selected model.*

Model	AIC	AIC weight
Survival ~ S + t + S*t + sex + τ + D + S*D + S* τ + sex*S + sex*t	1506.5	0
Survival ~ S + t + S*t + sex + τ + D + S*D + S* τ + sex*S	1488.6	0
Survival ~ S + t + S*t + sex + τ + D + S*D + S* τ	1486.6	0
Survival ~ S + t + S*t + sex + τ + D + S*D	1485	0
Survival ~ S + t + S*t + sex + D + S*D	1483	0.01
Survival ~ S + t + sex + D	1476	0.18
Survival ~ S + t + sex	**1474.1**	**0.47**
Survival ~ t + sex + D	1477.2	0.10
Survival ~ t + sex	1475.5	0.24

FIGURES

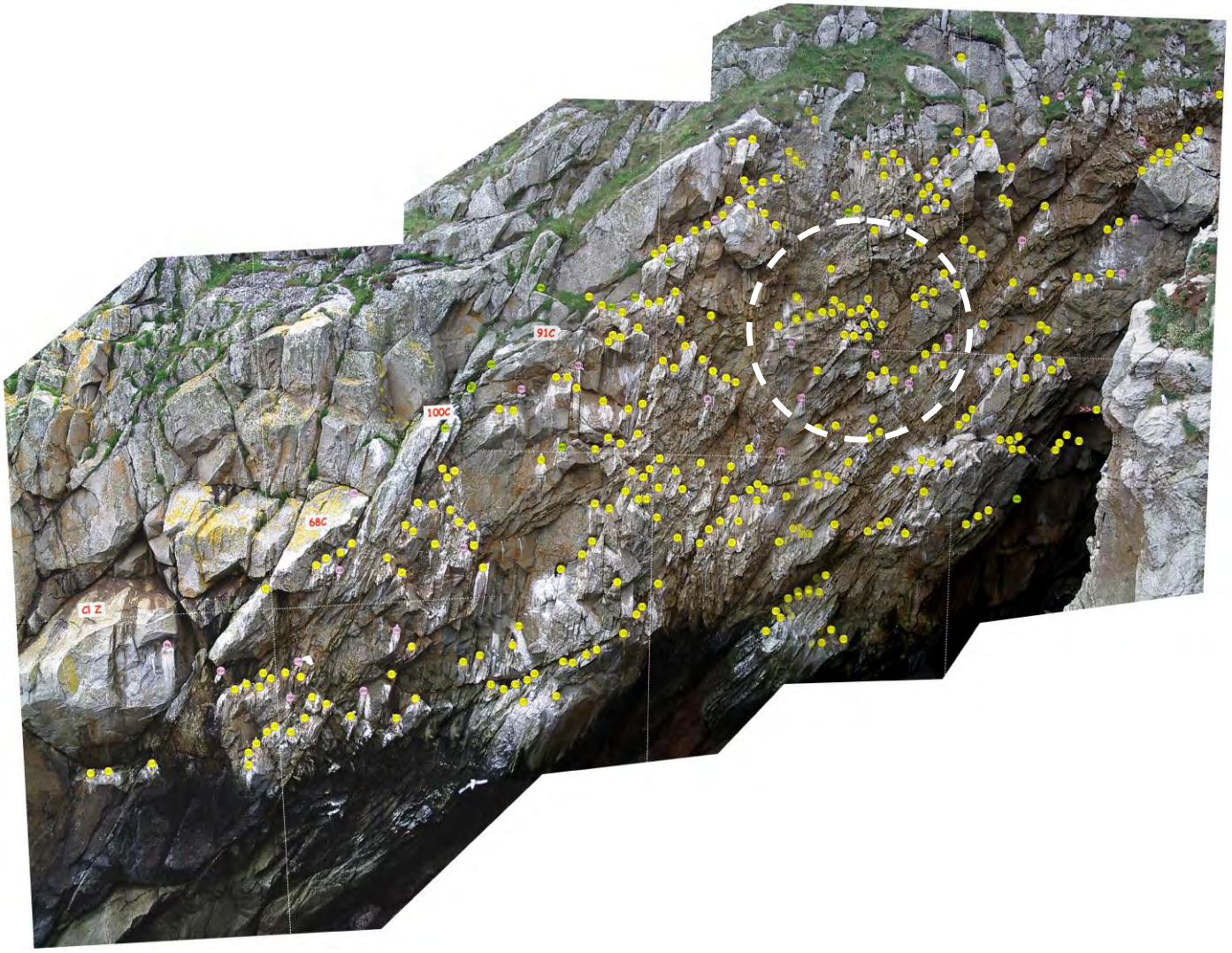
Figure 1. Portion of the studied cliff in a Cap-Sizun kittiwake colony. Some areas are so densely populated that nesting sites almost touch. Aggressive behavior between neighbors regularly occurs. (*photography by Gilles LeGuilloux*)

Figure 2. Map of the studied cliff and influence area of a nesting site. Each nesting site is represented by a dot. The white dashed circle indicates the influence area of one nesting site.

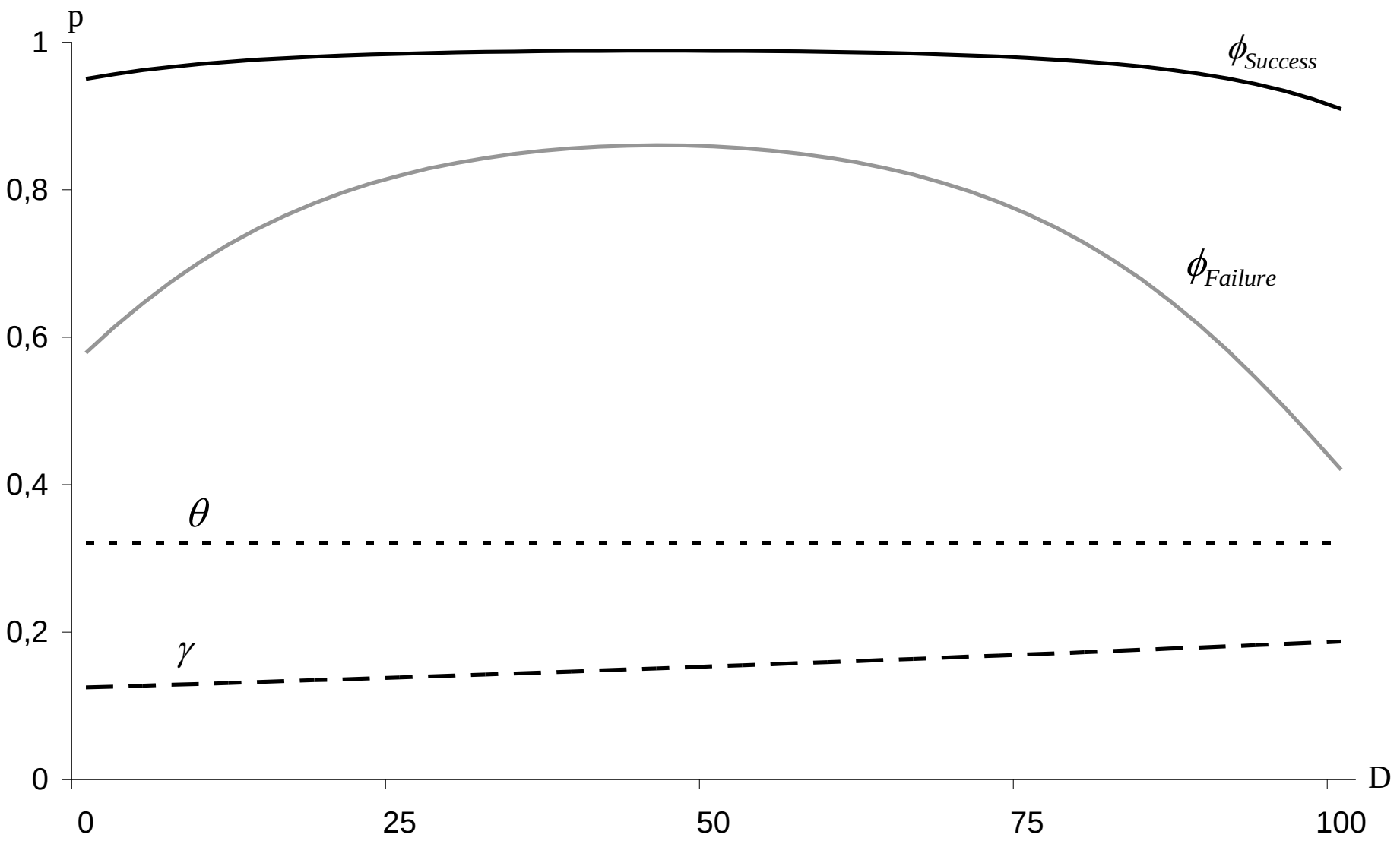
Figure 3. Estimates of Dynamics parameters as a function of local density *D* ('conspecifics presence'), for an average neighbor success rate. Black line: persistence probability of a successful site $\phi_{Success}$. Grey line: persistence probability of a failed site $\phi_{Failure}$. Dashed line: first colonization probability γ. Dotted line: recolonization probability θ.

Figure 4. Estimates of dynamics parameters as a function of neighbors' success rate $\tau_{i,t}$ ('public information'), for an average density. Black line: persistence probability of a successful site $\phi_{Success}$. Grey line: persistence probability of a failed site $\phi_{Failure}$. Dashed line: first colonization probability γ. Dotted line: recolonization probability θ.

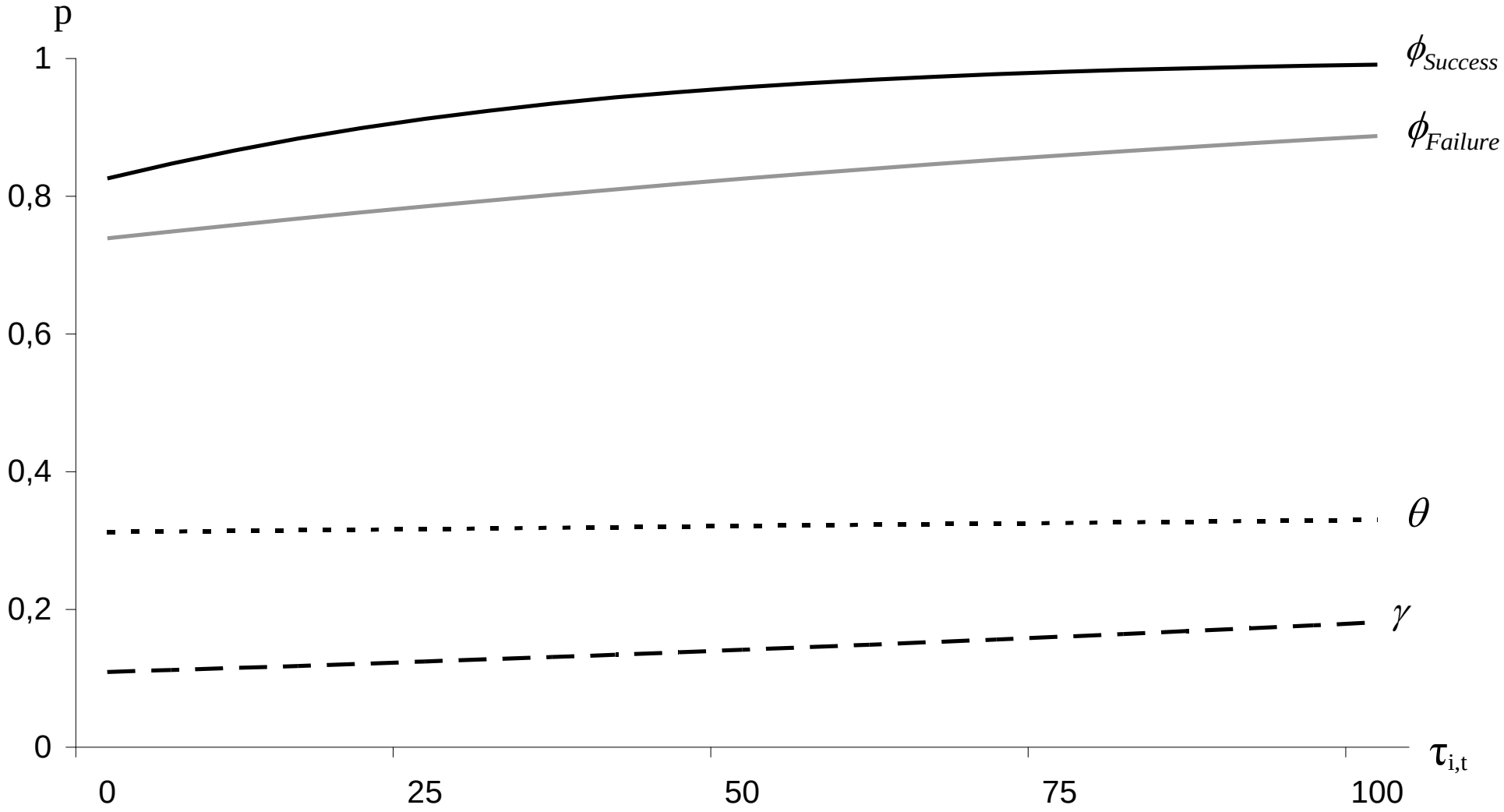
1



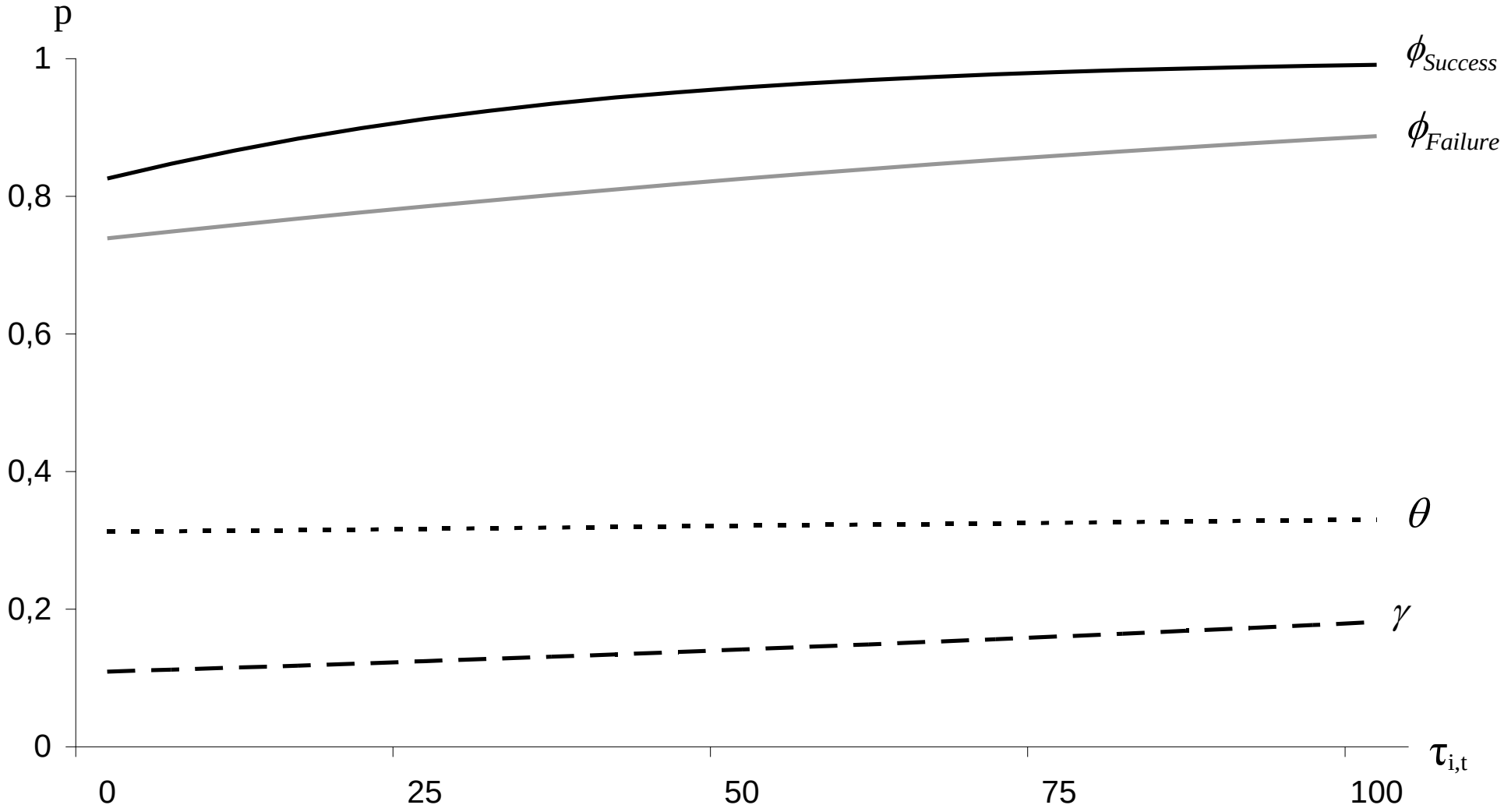
1


p
1
0,8
0,6
0,4
0,2
0
$\phi_{Success}$
$\phi_{Failure}$
θ
γ
D
0
25
50
75
100

1


p
1
0,8
0,6
0,4
0,2
0
0
25
50
75
100
$\phi_{Success}$
$\phi_{Failure}$
θ
γ
$\tau_{i,t}$

1


p
1
0,8
0,6
0,4
0,2
0
0
25
50
75
100
$\phi_{Success}$
$\phi_{Failure}$
θ
γ
$\tau_{i,t}$

1

CHAPTER IV

1

Estimating the occupancy rate of spatially rare or hard to detect species: a conditional approach

Jérôme A. Dupuis[1,*], Florent Bled[2], Jean Joachim[3]

[1]Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse, France

[2]Laboratoire Evolution et Diversité Biologique, Université Paul Sabatier, France

[3]INRA-CEFS, BP 27, 31326, Castanet-Tolosan, Toulouse, France

(*Biometrics, in press)*

Abstract – We consider the problem of estimating the occupancy rate of a target species in a region divided in spatial units (called quadrats); this quantity being defined as the proportion of quadrats occupied by this species. We mainly focus on spatially rare or hard to detect species which are typically detected in very few quadrats, and for which estimating the occupancy rate (with an acceptable precision) is problematical.

We develop a conditional approach for estimating the quantity of interest. The conditioning, which bears on the presence of the target species in the region of study, makes identifiable the occurrence and detectability parameters, regardless of the number of visits made in the sampled quadrats. Compared with an unconditional approach, it proves to be complementary, in that this allows us to deal with biological questions which cannot be addressed by the former.

Two bayesian analyses of the data are performed: one is non informative, and the other takes advantage of the fact that some prior information on detectability is available. It emerges that taking such a prior into account significantly improves the precision of the estimate when the target species has been detected in few quadrats and is known to be easily detectable.

Keywords: Bayesian estimation; Identifiability; Missing data; Occupancy rate; Spatially rare species.

*Corresponding author: dupuis@math.ups-tlse.fr

1. Introduction

The occupancy rate of a species in a specified region (called R) is a basic indicator of how the species in question occupies its habitat (e.g. MacKenzie *et al.*, 2002). The definition of this indicator presupposes that the region R is bounded and is composed of a finite number of spatial units. Biological motivations for considering this problem are various, but are often related to wildlife management concerns and endangered species conservation programs (Szaro & Johnston, 1996; Williams *et al.*, 2002). From a statistical point of view, we focus on species detected in very few quadrats, even not detected in any quadrats. Such data may occur with relatively high probability when the species in question is spatially rare or hard to detect (see Section 4.1.1). The statistical analysis of such sparse data by existing approaches yield significantly different estimations, while it is not the case if the target species has been well detected. This observation underlines the specificity of such data and stresses the necessity to understand why such distances occur. We below expound on the problematical, in details.

When modeling occupancy is of concern, Mackenzie *et al.* (2006) have developed an unconditional approach for estimating occupancy rate (see Section 4.5 of the book); their approach is unconditional in that it does not presuppose that the target species is present in the region R. From here on, it is important to point out that an unconditional approach involves two distinct quantities: one is the probability of presence of the species in question at a global scale (that is in the region R), the other is the probability of presence at a local scale (that is at the quadrat scale). Dupuis & Joachim (2006) have also developed an unconditional approach to model occurrence of species in the quadrats, but it is different from the one of MacKenzie. Both identically model the way a species present in R occupies the quadrats, but strongly differ regarding the way to model the presence at the global scale (details appear in Section 3.3.3). Therefore, it is expected that these two estimations differ when the target species has not been detected (which actually occurs), but are very close in the opposite case. Now, we have observed that they are still significantly different when it has been detected in very few quadrats. That means that, in such circumstances, the way one models the presence at the global scale may have a non negligible impact on the occupancy rate estimation of a detected species. The idea is to develop an alternative approach in which only the local scale is involved when the estimation concerns a detected species. Working conditionally on the presence of the target species in R is clearly a suitable strategy to achieve this objective. Interestingly, when the target species has not been detected by the sampling, the conditional approach allows to investigate some questions of biological interest which cannot be

addressed by the unconditional one (of course, the interpretation of the occupancy rate under such a conditional approach will differ). Let us give one example. It may occur that the presence (in R) of an undetected species has been proved by some additional information; thus, the conditional approach constitutes the natural framework to take this extra information into account.

Spatially rare species are typically detected in few quadrats, and it is thus expected that the estimate of the occupancy rate of such species is not very precise; hence the necessity to develop procedures to increase its precision (MacKenzie *et al.*, 2005). These authors have examined this question upstream (that is before collecting the data), by computing, for example, the minimal number of visits to be made in the sampled quadrats to achieve a given precision; see MacKenzie *et al.* (2006). In this paper we tackle the problem downstream. Our approach assumes that some prior information on the detectability parameter is available (for some species, such as birds, it is often the case; see Section 5). It is in fact expected that taking such a prior into account may have a positive impact on the estimate. Indeed, assume that it is well known that the target species is easily detectable; intuitively, it is clear that incorporating such a prior should remove some of uncertainty about its presence in the quadrats where it has not been detected, and therefore improve the precision of the estimate. We examine this question in detail in the framework of our illustration (Section 5).

2. The experimental protocol and data description

We assume that the region of study is composed of J spatial units (it includes, in particular, the situation where the region in question has been divided in a finite number of units). In the literature such units are also called sites or quadrats; for convenience, we will use this latter term. In this paper we assume that quadrats are of equal area. A sample of T quadrats is taken, and the sampled quadrats are numbered from 1 to T. Draw is usually performed at random so as to have a sample representative of the whole region R. Sometimes, all the J quadrats are explored; it typically occurs when J is not too large (the methodology developed in this paper also applies to this case). Lastly, an experimenter visits K times each sampled quadrat and records which target species have been detected. The list of target species is denoted by S (of course the list can include only one species). Detections are typically based on visual or oral recognitions; we assume that species are correctly identified. When $K = 4$ and $T = 6$, a possible record for a species $s \in S$ is: $y_s = 300040$. Such a sequence means that species s has been detected in quadrat 1 during three visits, and detected

in quadrat 5 during each visit. Moreover, its presence has not been detected in quadrats $2,3,4,6$.

When $y_s = \vec{0}$ (where $\vec{0}$ denotes the null vector) the target species has not been detected, and one does not know if the target species is present or not in the study region. It is not reasonable to exclude a priori such a record, simply because it may occur with relatively high probability, when the target species is spatially rare or hard to detect; see Section 4.1.1. This is also the option adopted by MacKenzie *et al.* (2006). From a statistical point of view, the problem is to estimate the proportion of quadrats occupied by any target species s from the record y_s. Before providing the assumptions which characterize the different approaches (conditional and unconditional) it is necessary, at this stage of the paper, to clarify in which context(s) each can be used. Unconditional approaches can be used as well with undetected species as with detected species. The conditional approach deals with species present in R; it can thus be used with any detected species, since such a species is necessarily present in R. When the target species has not been detected the conditional approach can still be used, but one has to keep in mind that the estimation of its occupancy rate is conditional on its presence in R. In fact, it provides an answer to the following question. If the target species is present in R (which one in fact ignores in such a situation), what is the value of its occupancy rate ? Moreover, when the target species has not been detected, its presence can sometimes be proved by another source of information (cf Section 5); in such a situation, the conditional approach allows to take into account this extra information (precisely via the conditioning).

3. Modeling detectability and occupancy

Throughout the paper, $p(.)$ denotes a probability mass function (p.m.f.) and $\mathbf{I}_{(C)}$ represents an indicator function that takes the value 1 when the condition C is true and zero otherwise. To describe the probabilistic assumptions related to occupancy and detectability, it is convenient to introduce the different processes which underlie the data y_s. It also allows to introduce, in a natural way, the parameters of biological interest.

3.1 Underlying processes.

For $j = 1,\ldots,J$, we denote by z_{sj} the indicator of presence of species s in quadrat j; thus, $z_{sj} = 1$ if species s is present in quadrat j, and 0 otherwise. Moreover, we set

$z_s = (z_{s1},\ldots,,\ldots,z_{sJ})$ and $z = \{z_s; s \in S\}$. The occupancy rate of species s, denoted by γ_s, can be expressed as a function of the z_{sj}'s, since $\gamma_s = \frac{1}{J}\sum_{j=1}^{J} z_{sj}$. Note that $\gamma_s \in \{\frac{1}{J},\frac{2}{J},\ldots,1\}$ when one works conditionally, and that $\gamma_s \in \{0,\frac{1}{J},\ldots,1\}$ in the opposite case. For a species s present in a sampled quadrat j, we denote by $x_{sj} \in \{0,1,\ldots,K\}$ the number of times that species s has been detected in quadrat j during the K visits. Note that x_{sj} is thus defined only conditionally on $z_{sj} = 1$. Moreover, we set $\mathrm{x}_s = (x_{s1},\ldots,,\ldots,x_{sT})$ and $\mathrm{x} = \{\mathrm{x}_s; s \in S\}$. It is useful to clarify the links between y_s, x_s and z_s. It is clear that $(y_{sj} = k) \Leftrightarrow (z_{sj} = 1$ and $x_{sj} = k)$, when $1 \le k \le K$. Moreover, the event $(y_{sj} = 0)$ covers in fact two exclusive situations: either species s is present in quadrat j but has not been detected, or it is not present in quadrat j (and cannot have been detected). Formally, one has the equivalence $(y_{sj} = 0) \Leftrightarrow (= 1$ and $x_{sj} = 0)$ or $(z_{sj} = 0)$. Consequently, z_{sj} is not known when species s has not been detected in quadrat j; in such a situation, z_{sj} is said to be missing. Note that the whole vector z_s is in fact missing when $y_s = \vec{0}$.

3.2 Modeling detectability

Biological assumptions related to detections are supported by the random vector x_s.

Assumption A1. We assume that the probability of detecting species s in quadrat j does not depend on its (possible) detections in the other quadrats.

Assumption A2. We assume that $x_{sj} \mid z_{sj} = 1$: $Binomial(K, q_s)$, where q_s represents the probability of detecting species s in any quadrat j during any visit (given that it is present in quadrat j).

For simplicity, detectability parameters do not depend on quadrats. Assumptions A1 and A2 are standard; they are also present in MacKenzie *et al.* (2006). Finally, it is useful to introduce the parameter $\mu_s = \Pr(x_{sj} \ge 1 \mid z_{sj} = 1) = 1 - (1 - q_s)^K$ which represents the probability of detecting species s in quadrat j (conditionally on its presence).

3.3 Modeling occupancy.

Biological assumptions related to how species s occupies the quadrats are supported by the random vector z_s. Moreover, it is useful to introduce the indicator ξ_s equal to 1 if species is present in R and zero otherwise. Moreover we denote by λ_s the probability that species s is present in R.

3.3.1. The unconditional model of MacKenzie et al..

To model occurrence of the target species s in the J quadrats, MacKenzie *et al.* (2006) assume that the z_{sj}'s are independent outcomes of a Bernouilli random variable. Hence, using the notation of MacKenzie for the occurrence parameter, one has:

$$p(z_s \mid \psi_s) = \psi_s^{N_s} (1-\psi_s)^{J-N_s} \tag{3.1}$$

where $N_s = \sum_{j=1}^{J} z_{sj}$ represents the number of quadrats in which species s is present, and $\psi_s = \Pr(z_{sj} = 1)$ (in the literature, ψ_s is also called occupancy parameter). Let us mention the paper of Royle & Kéry (2007) which provides an interesting discussion on the distinction to be made (from a biological point of view) between the occupancy rate γ_s and the occurrence parameter ψ_s. The MacKenzie *et al.* model gives a positive probability to the event $\xi_s = 0$, namely $(1-\psi_s)^J$; in that, their approach is unconditional (cf the introduction). Moreover, the above independence assumption implies that λ_s and ψ_s are linked since $\lambda_s = 1-(1-\psi_s)^J$. In other terms, the probability of presence at the global scale (λ_s) is completely determined by the one at the local scale (ψ_s); and vice versa, since one has: $\psi_s = 1-(1-\lambda_s)^{1/J}$.

3.3.2. The unconditional model of Dupuis and Joachim.

Dupuis & Joachim (2006) consider the hierarchical model below:

$$\xi_s \mid \lambda_s : \ Bernouilli(\lambda_s) \qquad p(z_s \mid \xi_s = 1, \varphi_s) = \frac{\varphi_s^{N_s}(1-\varphi_s)^{J-N_s}}{1-(1-\varphi_s)^J} \tag{3.2}$$

where the meaning of φ_s is given in the next Section (for convenience). Contrary to MacKenzie model, the z_{sj}'s are no more independent (details are omitted). More importantly, the unconditional model of Dupuis & Joachim (2006) involves two distinct parameters: namely λ_s (global scale) and φ_s (local scale), while in MacKenzie there is only a single parameter to deal with the two scales. In fact λ_s is here a nuisance parameter, that we do not

intend to estimate (contrary to Dupuis and Joachim, 2006); additional comments on this important point are provided at the end of the Section 4.2.

Compared with the Dupuis and Jochim approach, we observe that the MacKenzie one tends to promote *a priori* large values of λ_s. Indeed, if one places a uniform prior on ψ_s and one uses the fact that $\lambda_s = 1-(1-\psi_s)^J$, it is easy to check that $E(\lambda_s) = J/(J+1)$ which is close to 1. As an example: $\Pr(\lambda_s \geq 0.9) = 0.94$ if $J = 40$. On the contrary, if one places a uniform prior on λ_s, they are small values of ψ_s (and of γ_s, since $E(\gamma_s) = \psi_s$) which are now promoted *a priori*; indeed, one has $E(\psi_s) = 1/J+1$. We stress that, in the framework of this paper, we are in only concerned only by the first remark (since one assumes that the prior is placed on φ_s, not on λ_s). The situation is very different when one adopts the model of Dupuis and Jochim and uses non informative priors, since one has $E(\lambda_s) = 1/2$ and $E(\varphi_s) = 1/2$ (assuming of course that a uniform prior has been placed on these parameters). These remarks will be used to explain certain distances observed between the occupancy rate estimations yielded by these two approaches (cf Section 5.3.2).

3.3.3. The conditional model.

Let s be a species present in the study area R; thus $\xi_s = 1$, or equivalently $z_s \neq \vec{0}$. The p.m.f. (probability mass function) of z_s under the conditional model is:

$$p(z_s \mid \varphi_s) = \frac{\varphi_s^{N_s}(1-\varphi_s)^{J-N_s}}{1-(1-\varphi_s)^J} \tag{3.3}$$

where $\varphi_s \in]0,1[$. It is clear that if we start from (3.1) and compute the p.m.f. of $z_s \mid \xi_s = 1$, we obtain the p.m.f. of z_s under the conditional model. Moreover, it is easy to check that $\Pr(z_{sj} = 1 \mid \varphi_s) = \varphi_s/[1-(1-\varphi_s)^J]$ and that φ_s represents the probability that species s is present in quadrat j, given that it is present in at least one another quadrat (for the proof, see Web Appendix A). Note that ψ_s and φ_s do not have the same meaning, hence two distinct notations; in fact, φ_s is a conditional occurrence parameter (contrary to ψ_s). Note that γ_s is not a parameter of the model, contrary to φ_s; but both are linked, since $\mathrm{E}[\gamma_s \mid \varphi_s] = \mathrm{E}[z_{sj} \mid \varphi_s] = \Pr(z_{sj} = 1 \mid \varphi_s) = \varphi_s/[1-(1-\varphi_s)^J]$. Due to the conditioning on $\xi_s = 1$, the random variables z_{sj} are not independent (contrary to the MacKenzie model): for example, $\Pr(z_{sj} = 1 \mid \varphi_s) = \varphi_s/[1-(1-\varphi_s)^J]$, but $\Pr(z_{sj} = 1 \mid \varphi_s, \bar{z}_{sj} = \vec{0}) = 1$ where $\bar{z}_{sj}$ denotes the vector

$z_s \setminus \{z_{sj}\}$. However, a certain form of conditional independence between the z_{sj}'s holds, as stated by the proposition (3.1).

Proposition 3.1. Let i, j be two any quadrats, z_{si} and z_{sj} are independent, given that species s is present in at least one another quadrat (distinct from i and j).

Proof. See Web Appendix A.

From a biological point of view, we thus assume that the presence of species s in quadrat i is not affected by the presence of that species in quadrat j (given that it is present elsewhere); this assumption is standard (disregarding the conditioning).

The key difference between an unconditional approach and a conditional approach is that the latter models only the way a species present in R occupies the J quadrats (local stage), whereas the former deals with the two scales (local and global). It is important to point out that the three approaches identically model the way a species present in R occupies the J quadrats. Lastly, it is of interest to note that the two unconditional approaches lead to a same conditional version, while, from this latter, at least two unconditional versions can be built.

4. Likelihoods, identifiability, estimation and computational issues.

From now, the approach which assumes A1, A2 and models occupancy via (3.1) is called the unconditional approach of MacKenzie (or the MacKenzie approach, in shortened form). The one which models occupancy via (3.2) is called the unconditional approach of Dupuis & Joachim (or the Dupuis & Joachim approach), and the one which models occupancy via (3.3) is simply called the conditional approach (both assuming of course A1 and A2).

4.1 Likelihoods.

4.1.1 Conditional approach.

The parameter of the model is $\theta = \{\varphi_s, q_s; s \in S\}$. The likelihood of θ based on data $y = \{y_s; s \in S\}$ is: $L(\theta; y) = \prod_{s \in S} p(y_s \mid \varphi_s, q_s)$. Note that we thus implicitly assume that species behave independently with respect to occupancy in the J quadrats. For any species s such that $y_s \neq \vec{0}$, we have:

$$p(y_s \mid \varphi_s, q_s) = \frac{\rho_s \varphi_s^{V_s} q_s^{W_s} [1-q_s]^{U_s} [(1-q_s)^K \varphi_s + 1 - \varphi_s]^{T-V_s}}{1-(1-\varphi_s)^J} \qquad (4.1)$$

where V_s denotes the number of quadrats in which species s has been detected, W_s the total number of visits during which species s has been detected, $U_s = KV_s - W_s$, and

$$\rho_s = \prod_{j=1}^{T} \binom{K}{y_{sj}};$$

moreover, we have:

$$p(y_s \mid \varphi_s, q_s) = \frac{((1-q_s)^K \varphi_s + 1 - \varphi_s)^T - (1-\varphi_s)^J}{1-(1-\varphi_s)^J} \qquad (4.2)$$

if $y_s = \vec{0}$. Computational details involve rather lengthy developments and have been omitted for brevity, but they appear in a technical report available from the first author on request. The above formula also applies in the particular case $K = 1$. The formulae (4.1) and (4.2) allow to compute (as a function of J, T, K, φ_s and q_s) the probabilities that a species s present in R has not been detected (or has been very few detected); results underline that such data occur with relatively high probabilities when species s is spatially rare (as stressed in the introduction). Let us give two examples (in Web Appendix B, we provide other examples which are commented in detail). Let $p_0 = \Pr[(V_s, W_s) = (0,0)]$ and $p_1 = \Pr[(V_s, W_s) = (1,1)]$. One has p_0=0.20, p_1=0.29, if J=T=20, K=4, $\varphi_s = 0.1$, $q_s = 0.2$ (that is $\mu_s \approx 0.60$); and p0=0.38, p1=0.23, if J=40, T=20, K=4, $\varphi_s = 0.05$, $q_s = 0.3$ (that is $\mu_s \approx 0.75$); where $\mu_s = 1-(1-q_s)^K$ (cf Section 3.2).

4.1.2 Unconditional approaches.

- The MacKenzie approach.

The associated model is parametrized by $\{\psi_s, q_s; s \in S\}$, and one has:

$$p(y_s \mid \psi_s, q_s) = \rho_s \psi_s^{V_s} q_s^{W_s} [1-q_s]^{U_s} [(1-q_s)^K \psi_s + 1 - \psi_s]^{T-V_s} \qquad (4.3)$$

from which we deduce the likelihood. Note, that contrary to the conditional approach, J plays no part in the MacKenzie likelihood.

- The Dupuis & Joachim approach.

We stress that the result below is new. Indeed, in Dupuis & Joachim (2006), the aim was to estimate the species richness of a biological community from quadrat sampling data, not occupancy rate. The expression of $p(y_s \mid \varphi_s, q_s, \lambda_s)$ below is derived from (4.1) and (4.2); details are omitted. We have:

$$p(y_s \mid \varphi_s, q_s, \lambda_s) = \frac{((1-q_s)^K \varphi_s + 1 - \varphi_s)^T - (1-\varphi_s)^J}{1-(1-\varphi_s)^J \lambda_s + (1-\lambda_s)} \tag{4.4}$$

if $y_s = \vec{0}$ and

$$p(y_s \mid \varphi_s, q_s, \lambda_s) = \frac{\rho_s \varphi_s^{V_s} q_s^{W_s} [1-q_s]^{U_s} [(1-q_s)^K \varphi_s + 1 - \varphi_s]^{T-V_s}}{1-(1-\varphi_s)^J \lambda_s} \tag{4.5}$$

otherwise. It is clear that, when $y_s \neq \vec{0}$, the estimations of φ_s, q_s and γ_s based on (4.5) coincide with those yielded by the conditional approach.

4.2 Identifiability issues.

- Recall that the parameters φ_s and q_s of the MacKenzie approach are not identifiable when the sampled quadrats are visited only once. Indeed, when $K = 1$, the parameters φ_s and q_s appears in the MacKenzie likelihood only through the product $\varphi_s q_s$.
- When we use the conditional approach, we have the following result.

Proposition 4.1 The model associated with the conditional model is identifiable, regardless of the number K of visits in the sampled quadrats. In particular, the parameters q_s and φ_s are identifiable when the the sampled quadrats are visited only once.

Proof. See Web Appendix C.

Surprisingly, the model is indeed identifiable when $K = 1$. This means that data collected via this protocol, improperly called presence-absence data and afterwards called $1-0$ data, can theoretically be used to estimate φ_s, γ_s, and q_s ($= \mu_s$ when $K = 1$). However, we do not recommend the use of $1-0$ data to estimate occupancy rates, except when some prior information on q_s is available. We have indeed observed (in the framework of our illustration) that estimations of γ_s based on $1-0$ data were very close to those based on data with multiple visits, in the presence of prior information on q_s. Of course, in this comparison, we assume that the sampling effort is constant, in that μ_s takes the same value under both situations . On the contrary, when a flat prior is placed on q_s (and on φ_s), we have observed that estimations of q_s and φ_s were very close, which is clearly the sign that parameters are badly estimated.

- The following proposition indicates the condition on K on which the model based on the unconditional approach of Dupuis & Joachim is identifiable.

Proposition 4.2 The model based on the Dupuis & Joachim approach is identifiable, whatever the value of $K \geq 2$, but it is not identifiable if $K = 1$.

Proof. See Web Appendix D. Only the first part of the result is new; indeed, Dupuis & Joachim (2006) have already mentioned the absence of identifiability when $K = 1$.

It is clear that the parameter λ_s (though identifiable when $K \geq 2$) cannot be estimated, in a useful way, from a single record y_s. Estimating them with a reasonable precision will actually require several datasets of y_s; see Dupuis & Joachim (2006). In our framework, this is not a problem, since estimating γ_s is only of interest. The parameter λ_s is here a nuisance parameter, which has yet to be introduced in the model when one wishes to adopt an unconditional approach and to model separately the presence of the target species at the local and global scales (cf Section 3.3.2).

4.3 Estimation and computational issues

Obtaining the bayesian estimations of the quantities of interest requires the use of Markov Chain Monte Carlo methods (MCMC). In this Section we provide the main outline of the MCMC algorithm to be implemented for the conditional approach. Since we want to estimate both γ_s (which is a function of z_s), and parameters q_s and φ_s, we implement a MCMC algorithm which takes advantage of the missing data structure mentioned in Section 3.1. The MCMC algorithm is thus implemented on (θ, z_m) where $z_m = \{z_s^{mis}; s \in S\}$ and z_s^{mis} denotes the set of the missing z_{sj}'s related to s. Such a strategy is standard in missing data models for estimating the parameters or any function of z_m; see e.g. Robert & Casella (1999), or Dupuis (1995) in a capture-recapture set-up. Before indicating how to implement these two steps, we need to provide the expression of the complete data likelihood denoted by $L(\theta; y, z_m)$. By using A1, A2 and (3.3), and by observing that (y, z_m) and (x, z) provide the same information on θ, we have:

$$L(\theta; y, z_m) = p(\mathrm{x}, z \mid \theta) = \prod_{s \in S} p(\mathrm{x}_s, z_s \mid q_s, \varphi_s)$$

where

$$p(\mathrm{x}_s, z_s \mid q_s, \varphi_s) = \frac{\rho_s q_s^{W_s} (1-q_s)^{Kn_s - W_s} \varphi_s^{N_s} (1-\varphi_s)^{J-N_s}}{1-(1-\varphi_s)^J} \qquad (4.6)$$

where n_s represents the number of sampled quadrats in which species s is present.

Updating the parameters proceeds as follows: φ_s is updated via a Metropolis-Hastings step; and q_s via a Gibbs step: q_s : Beta $(1+W_s, 1+Kn_s - W_s)$ (assuming that a uniform distribution is placed on q_s). Implementing the augmentation data step requires a particular attention. Details appear in Web Appendix E. Lastly, the bayesian estimate of γ_s is easily obtained by applying the ergodic theorem (details are omitted).

The MCMC algorithm to be implemented for the unconditional approach of Dupuis and Joachim, is described in Web Appendix F.

5. An illustration

The goal in this Section is to illustrate from real data the different points examined in the previous Sections, not to investigate a particular biological question. The list of chosen species has been made up with this aim.

5.1 Data description

Data considered in this paper are a part of a more important data set collected in May 1985 to estimate the number of nesting bird species present in the forest of Montech (located near Toulouse in France). The protocol and the field description are described in details in Dupuis & Joachim (2006) who have already analysed data from the forest of Montech, but collected in 1987 (not in 1985); therefore only the main points are given here. This forest, with a surface area of 1000 hectares, is relatively spatially homogeneous (mainly composed of oaks and hornbeams). It has been divided into 40 quadrats; 22 quadrats are located in the inner forest, and 18 in outer forest. All the quadrats have been visited. Our objective is, for each species belonging to the list S given further, to estimate its occupancy rate in the inner forest (thus $T = J = 22$). We consider that species s occupies quadrat j if at least one individual belonging to species s has nested in quadrat j, during may 1985. Information about the presence of nesting species was provided by acoustic recognition of singing males according to the following procedure. The researcher spent a prescribed time (twenty minutes in our study) at each station (in the center of quadrat), listening to birds. More precisely, data have been collected according to the following point count protocol: each 20-minute session has been sliced in four subsessions of 5 minutes each, during which the experimenter records whether the presence of the species of interest has been detected, or not. Each slice is the

equivalent of a visit, therefore $K = 4$. Table 1 below provides the data set; more precisely, for each selected species s, we have indicated the values of the sufficient statistics (V_s, W_s).

Table 1. The data set: number of quadrats in which species *s* has been detected (V_s) and number of visits during which species s has been detected (W_s)

s	Bullfinch	Marsh tit	Coal tit	Firecrest	Greenfinch	Wren	Black bird
(V_s, W_s)	(0,0)	(0,0)	(0,0)	(1,1)	(1,1)	(8,14)	(17,32)

The list S includes seven species, namely: the bullfinch, the coal tit, the marsh tit, the firecrest, the greenfinch, the wren and the black bird. Three species (the bullfinch, the coal tit and the marsh tit) have not been detected by the quadrat sampling. Interestingly, the presence of the marsh tit has been proved by another source of information (mistnet capture for ringing and monitoring local birds). The greenfinch and the firecrest have been detected in only one quadrat (and during only one visit), the black bird has been detected in most quadrats, and the situation of the wren is intermediate.

5.2 Prior information

The data set analyzed in this paper is a part of broad study of birds populations of Midi-Pyrénées (which is a region located in the west-south of France). This study started around 1985; consequently; prior information is now relatively abundant. However, we have not used the totality of this information. Indeed, we have taken into account only the information related to q_s; the prior information available on $\varphi_{s\ s}$ having been ignored, considering that this parameter was too directly linked to the quantity of interest γ_s.

Information available in the literature about how much a bird species is detectable typically consists in providing its level of detectability, namely: discrete (group 1), moderately detectable (group 2), or easily detectable (group 3). During the breeding period, the bullfinch and the firecrest belong to the group 1, the marsh tit to the group 2, and the coal tit, the greenfinch, the black bird to the group 3 (Blotzheim & Bauer, 1985, 1988, 1993, 1997). In Dupuis & Joachim (2006) we have determined beta distributions for each group (from external data). Our intention is to use these prior distributions, but the difficulty is that Dupuis & Joachim (2006) worked with the probability of detecting species during the whole session, that is μ_s, not q_s. To deduce the beta distribution on q_s (from the one placed on μ_s) we have

simply used classical Monte Carlo methods, by observing that $q_s = 1-(1-\mu_s)^{1/K}$. Note that we implicitly assume that the distribution of q_s is well approximated by a Beta distribution (it is effectively the case in our context).

Table 2. Prior information on detectability: coefficients of the Beta distribution placed on μ_s, prior mean and 95% credible interval of μ_s, and coefficients of the Beta distribution placed on q_s

s	Bullfinch	Marsh tit	Coal tit	Firecrest	Greenfinch	Wren	Black bird
μ_s	$(4,6)$	$(9,6)$	$(16,4)$	$(4,6)$	$(16,4)$	$(16,4)$	$(16,4)$
mean	0.4	0.6	0.8	0.4	0.8	0.8	0.8
C.I.	$[0.05,0.65]$	$[0.3,0.8]$	$[0.65,1]$	$[0.05,0.65]$	$[0.65,1]$	$[0.65,1]$	$[0.65,1]$
q_s	$(3.9,27.3)$	$(8.4,31.2)$	$(12.9,24.6)$	$(3.9,27.3)$	$(12.9,24.6)$	$(12.9,24.6)$	$(12.9,24.6)$

Table 2 above provides for each species $s \in S$, the prior distribution on μ_s, as well the resulting prior mean and a 95% credible interval. It clearly appears that the prior used is relatively moderate (see the length of the 95% prior credible intervals). Table 2 also provides the beta distribution placed on each q_s.

5.3 Results

This Section is organized as follows. In the first part we provide the estimations of γ_s (for each species $s \in S$) yielded by the conditional approach. Then we provide estimations yielded by the two unconditional approaches.

5.3.1 By using the conditional approach.

For each species $s \in S$, Table 3 below provides the posterior mean and 95% credible interval of γ_s, under informative and non informative priors. (The value 0.05 which appears in a certain number of intervals is the approximated value of $1/J$.) For the marsh tit, the conditional approach allows to take into account the fact it was present in the region of study in 1985 (whereas it has not been detected by the quadrat sampling). The bullfinch has not been detected by the sampling and we do not have some any extra information that will verify its presence (contrary to the marsh tit). The value appearing in Table 3 thus represents the (estimated) occupancy rate of the bullfinch, if it was present in the forest of Montech during spring 1985 (cf Section 2). This remark applies also to the coal tit.

Table 3. Bayesian estimations of occupancy rates under the conditional approach: posterior means and 95% posterior credible intervals, with and without prior information on detectability.

s	Bullfinch	Marsh tit	Coal tit	Firecrest	Greenfinch	Wren	Black bird
without prior	0.25 [0.05, 0.82]	0.25 [0.05, 0.82]	0.25 [0.05, 0.82]	0.25 [0.05, 0.82]	0.25 [0.05, 0.82]	0.47 [0.36, 0.68]	0.87 [0.77,1]
with prior	0.15 [0.05, 0.50]	0.08 [0.05, 0.23]	0.06 [0.05, 0.14]	0.24 [0.05, 0.73]	0.07 [0.05, 0.14]	0.47 [0.36, 0.64]	0.89 [0.77,1]

We now comment on the results in Table 3.

• Concerning the undetected species (that is the bullfinch, the marsh tit and the coal tit), we note that, in the absence of prior information, the bayesian estimate of the occupancy rate is imprecise (the range of the posterior 95% credible interval being particularly large). This is why we provide the posterior distribution $N_s = J\gamma_s$ (in fact the histogram of the MCMC sequence produced by Matlab) which gives more information than the mean and the 95% credible interval. (The histogram related to γ_s is less legible, hence this choice.)

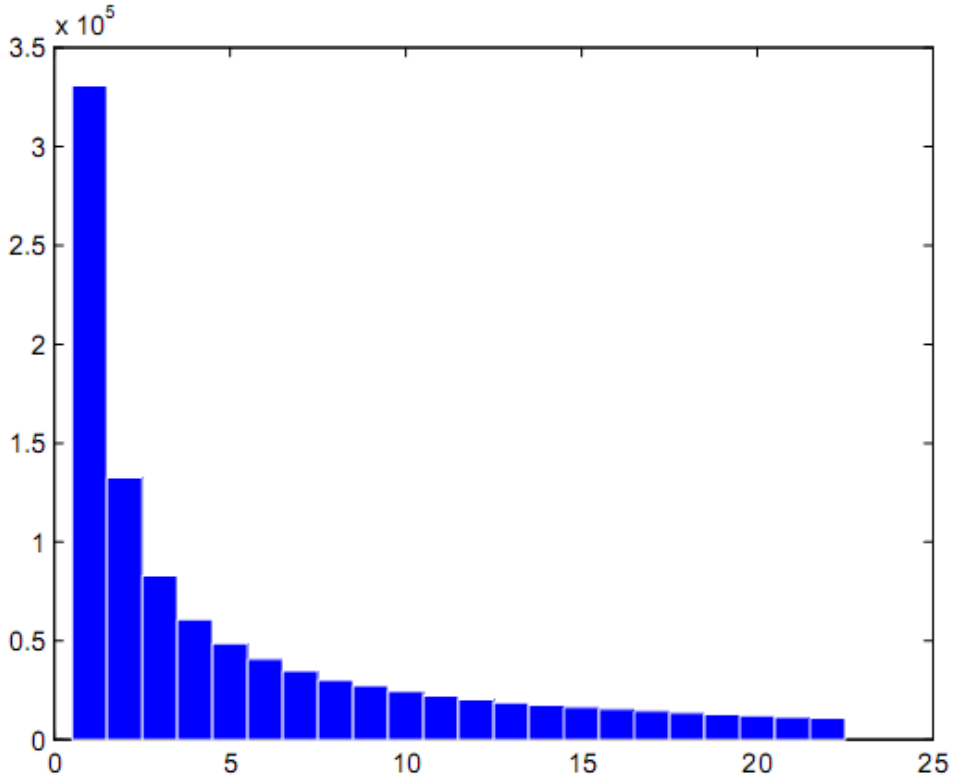


Figure 1. Histogram of the MCMC sequence related to $N_s = J\gamma_s$ (L=10^6 iterations) for the marsh tit: the non informative case.

The histogram (Figure 1) points out the presence of a mode strongly marked around $N_s = 1$; information which is completely missed when one retains the mean and 95% credible interval as a summary of the posterior distribution. Taking into account the prior information available on detectability significantly improves the precision of estimations. It is particularly marked for the coal tit and the marsh tit. Not surprisingly, we observe that the more a species is detectable, the greater the precision of the posterior estimates. Similarly, specifying a more informative prior on the parameters also improves the precision of the posterior estimates. It is of interest to examine how the posterior distribution of N_s is modified by the prior and to compare both posteriors (with and without prior). It is why we provide, for the marsh tit (as an example), the histogram of the MCMC sequence related to N_s (cf Figure 2).

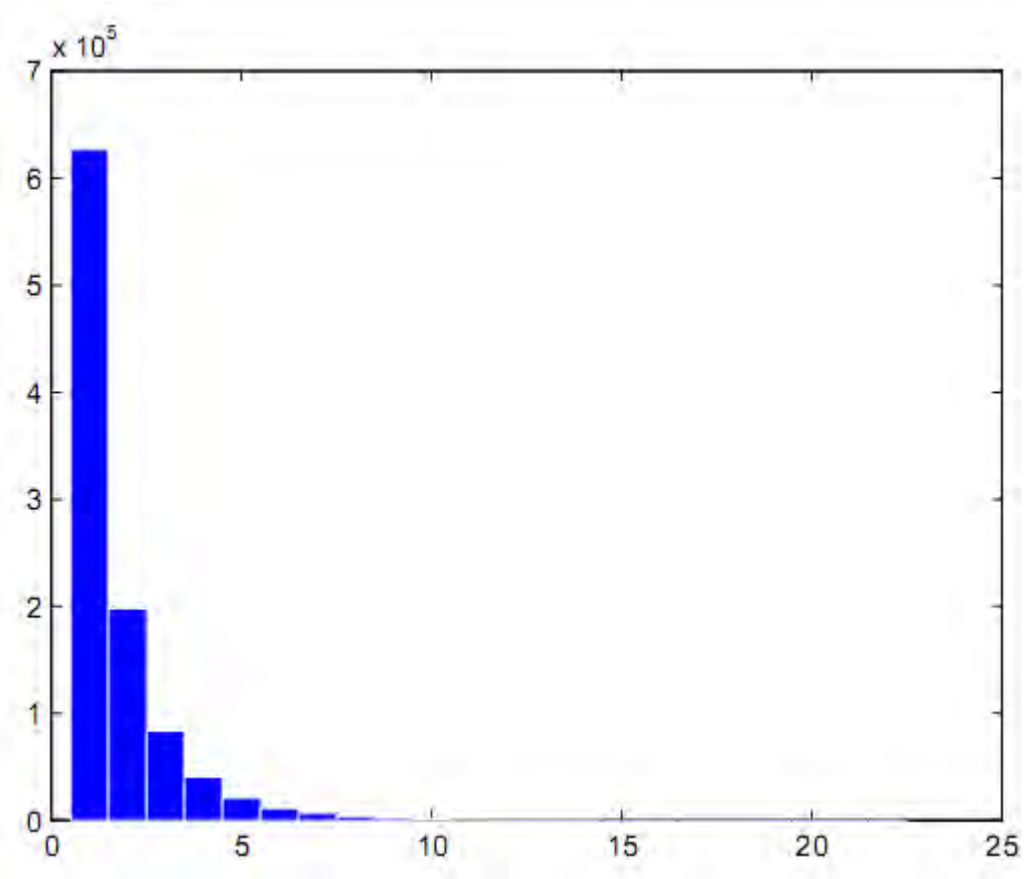


Figure 2. Histogram of the MCMC sequence related to $N_s = J\gamma_s$ (L=10^6 iterations) for the marsh tit: the informative case.

We observe that the shape of the posterior distribution is very similar, simply the essential of the probability mass is now concentrated in few values (close to one). Note that the posterior mode is not changed.

- Concerning the species detected in only one quadrat (that is the firecrest and the greenfinch) very similar comments can be made on the non informative bayesian estimation of γ_s (see above). We observe that taking into account the prior significantly increases the precision of the estimation for the greenfinch (which is easily detectable). Conversely, the improvement is not so marked for the firecrest (which is relatively discrete), and we note that the posterior mean of γ_s is very little modified by the (informative) prior.
- The wren has been detected in about 1/3 of the quadrats, and, not surprisingly, the non informative estimation of γ_s is much more precise than the three previous ones. Moreover, we observe that the impact of the (informative) prior is minute. The black bird has been detected in a large number of quadrats. Not surprisingly, the estimation of γ_s is quite precise, even in the absence of prior information; and taking into account the prior has little effect on estimation.

5.3.2 By using an unconditional approach.

We provide in Table 4 the posterior mean and 95% credible interval of γ_s yielded by the unconditional approaches. Estimations appearing in Table 4 have been computed by placing a uniform distribution on each parameter (namely on q_s, ψ_s, φ_s, as well on λ_s when the unconditional approach of Dupuis & Joachim has been used). We limited ourselves to

Table 4. Non informative bayesian estimations of occupancy rates (posterior means and 95% posterior credible intervals) yielded by the unconditional approaches of MacKenzie and Dupuis & Joachim, and by the conditional approach

Species	MacKenzie	Dupuis & Joachim	Conditional
Marsh tit, Bullfinch,	0.13	0.01	0.25
Coal tit	[0, 0.73]	[0, 0.09]	[0.05, 0.82]
Firecrest,, Greenfinch	0.29	0.25	0.25
	[0.05, 0.86]	[0.05, 0.82]	[0.05, 0.82]

undetected species (bullfinch, marsh tit, and coal tit) and to species detected during only one visit (greenfinch and firecrest); estimations of γ_s related to well detected species (wren and black bird) do not appear in Table 4, simply because they are very close to those yielded by

the conditional approach (see Table 4). Results appearing in Table 4 have been gathered in two groups: undetected species and the others (since estimations are the same within a group). For all these species, we have again indicated the estimation furnished by the conditional approach, to facilitate comparisons.

We begin by comparing the results provided by the unconditional approaches. For the species detected during only one visit (greenfinch and firecrest), the distance between both estimations is not negligible: 0.29 for the MacKenzie approach, and 0.25 for the one of Dupuis & Joachim. For the undetected species (bullfinch, marsh tit, and coal tit), we observe important distances between: 0.13 for MacKenzie, 0.01 for Dupuis & Joachim. We explain these distances by the fact these two approaches strongly differ regarding the way the presence at the global scale is modelled (cf Section 3.3.2). Recall indeed that the MacKenzie approach promotes *a priori* the high values of λ_s, while $\mathrm{E}[\lambda_s] = 1/2$ under the Dupuis & Joachim approach. Consequently, it is not surprising that the posterior mean of γ_s is significantly smaller under the Dupuis & Joachim approach, since γ_s and λ_s are linked by the relation $\Pr(\gamma_s = 0) = 1 - \lambda_s$. We observe that the most important distance occurs when $y_s = \vec{0}$; therefore (and not surprisingly) the way one models occupancy at the global scale plays a particularly important part in the estimation of γ_s, mainly when species *s* has not been detected.

Now, it is interesting to compare the occupancy rate estimation of the marsh tit computed under the conditional approach, with the one computed under the unconditional approach of Dupuis & Joachim. Taking into account (via the conditional approach) the fact that the marsh tit was present in the forest in 1985 strongly modifies the estimate of the occupancy rate provided by the unconditional approach: 0.01 for the latter vs 0.25 for the former. We explain this important distance as follows. Under the unconditional approach the probability $\Pr(\xi_s = 1 \mid y_s = \vec{0})$ is low, namely 0.04 (value obtained by using the MCMC algorithm described in Section 4.3) and $\Pr(\gamma_s = 0 \mid y_s = \vec{0}) = 1 - \Pr(\xi_s = 1 \mid y_s = \vec{0}) = 0.96$ is thus strong, while it is equal to zero in the conditional approach. Hence the small value of $\mathrm{E}[\gamma_s \mid y_s]$ under the unconditional approach and the important distance between both estimations. As information, the informative bayesian estimation of γ_s under the unconditional approach is equal to 0.007 (while it is equal to 0.08 under the conditional one, cf Table 3).

Note that the estimations of γ_s under the Dupuis & Joachim and conditional approaches coincide when the species has been detected during only one visit (firecrest and greenfinch); this is in fact a general result (cf Section 4.1.2). On the other hand, we stress that, under the conditional approach, this is not true when a species is undetected; it is in fact slightly smaller (the distance is around 0.004); values appearing in Table 4 have been rounded off.

6. Discussion and conclusion

We have developed a conditional approach for estimating the occupancy rate of a target species in a region *R* composed of *J* quadrats; it provides an alternative to the existing approaches which are unconditional: namely, the one expounded in the book of MacKenzie *et al.* (2006), and the other based on the occupancy model considered in Dupuis & Joachim (2006). In the framework of spatially rare or hard to detect species (which are typically detected in very few quadrats), the conditional approach presents two advantages: first, it does not require the modeling of the presence of the target species at the global scale, that is in the region *R* (contrary to an unconditional approach); second, when the species of interest has not been detected by the sampling, this allows us to deal with specific questions which cannot be addressed by an unconditional approach. Moreover, though we have motivated the conditional approach from species detected in very few quadrats, it is clear that it is also interesting, from a conceptual point of view, outside this specific framework. The occupancy rate estimations yielded by these different approaches are either identical or very close when the target species has been well detected. In the opposite case (which constitutes here the situation of main interest), the results can differ significantly, though not necessarily. We now comment on these results and provide some guidelines to indicate which approach is the most appropriate considering the question of interest.

- We begin by comparing the two unconditional approaches. We have observed that they produce significantly different estimations only when the target species has been few observations, or not even detected. We consider that it is the way one models occupancy at the global scale which is responsible of such distances. Recall that in MacKenzie *et al.* (2006) the probability of presence at the global scale (λ_s) is completely determined by the one at the local scale (ψ_s), and vice versa; by contrast, Dupuis & Joachim have two independent parameters, one by scale. We thus believe that the MacKenzie approach is the one to be used

when the link between λ_s and ψ_s seems reasonable, while the second one is suitable in the absence of any information about a possible link between these two quantities.

- We now compare the unconditional approach of Dupuis & Joachim and the conditional one. Firstly, both coincide when the target species has been detected (either frequently or infrequently). The Dupuis & Joachim approach exhibits a drawback: it introduces a nuisance parameter (λ_s) that is clearly unnecessary when estimation concerns a detected species. Consequently, we consider that, in such circumstances, the conditional approach is conceptually preferable, since it is more parsimonious. Nevertheless, we stress that the latter cannot substitute to the former, when the target species has not been detected by the sampling. Recall that, in such circumstances, the occupancy rate has two distinct interpretations, both of which are useful from a biological point of view. Lastly, when the target species has not been detected by the sampling, the researcher may sometimes have some prior information on its probability of presence in the region R; it is interesting to point out that the unconditional approach of Dupuis & Joachim will thus be particularly appropriate to deal with such a situation, by adopting a prior on the parameter λ_s that will reflect this opinion.

Assume that the target species has not been detected by the sampling and that its presence in R has been proved by some additional information. In our mind, this additional information is limited to the knowledge that the target species is present in R; but, as pointed out by one Referee, the data on which this extra information is based, can sometimes be available. He has thus suggested (rightly) that the conditional approach should be modified to take it into account in the model. Below we indicate how to proceed. For brevity, we limit ourselves to the situation where the structure of the auxiliary data (denoted by y'_s) is similar to the one of y_s (but the sampled units used to collect y'_s may be different from those used for y_s). The probability mass function of y'_s will thus be typically parametrized by φ_s and q'_s (where the latter is a priori distinct from q_s since the way of obtaining information contained in y'_s can differ from the one of y_s). The MCMC algorithm implemented in our paper can be easily modified to deal with such a situation. For example, if, for both data y_s and y'_s, the quadrat j has been sampled and is such that $y_{sj} = 0$, thus we simulate the missing z_{sj} according to a Bernoulli with parameter

$$Pr(z_{sj} = 1 \mid y_{sj} = 0, y'_{sj} = 0, \varphi_s, q_s, q'_s) \propto (1-q_s)(1-q'_s)\frac{\varphi_s}{1-(1-_s)^J},$$

by assuming that y_{sj} and y'_{sj} are independent conditionally on z_{sj}.

When a species s has been detected in a small number of quadrats the precision of the non informative bayesian estimate of γ_s is low. Incorporating some prior information on q_s - assuming of course that such a prior is available - is an efficient strategy when the target species is known to be easily detectable, as illustrated in this paper. Though the strategy of MacKenzie *et al.* (2006) is completely different (see the introduction), it is of interest to observe that both actually rely on the same general idea explained in MacKenzie *et al.* (2005), namely borrowing information on detectability from different sources to increase the precision of the occupancy rate estimates.

The models considered in this paper are simple, but well suitable for our purpose: introducing the conditional approach and comparing it to the existing approaches. Of course, more complex versions of the conditional approach will have to be developed in future, to deal (for instance) with some possible spatial heterogeneity at the quadrat scale.

Acknowledgements

The authors are grateful to Emmanuelle Cam for helpful comments.

References

Blotzheim U.N. and Bauer K.M. (1985, 1988, 1993, 1997). *Handbuch der Vogel Mitteleuropas*. Aula-Verlag, Wiesbaden.

Dupuis, J. A. and Joachim, J. (2006) Bayesian estimation of species richness from quadrat sampling data in the presence of prior information. *Biometrics* **62**, 706-712.

Dupuis, J. A. (1995) Bayesian estimation of movement and survival probabilities from capture-recapture data. *Biometrika* **82**, 761-772.

MacKenzie, D.I., Nichols, J.D., Royle J.A., Pollock, K.H., Bailey, L.L., Hines, J.E. (2006) *Occupancy Estimation and Modeling*. Elsevier, Amsterdam.

MacKenzie, D.I., Nichols, J.D., Sutton N., Kawanishi K., Bailey, L.L. (2005) Improving inferences in population studies of rare species that are detected imperfectly. *Ecology* **86**, 1101-1013.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle A.J. and Langtimm, A.C. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**, 2248-2255.

Robert, C.P. and Casella,G. (1999). *Monte-Carlo Statistical Methods*. Springer-Verlag.

Royle, J.A. and Kéry, M. (2007). A bayesian state-space formulation of dynamic occupancy models. *Ecology* **88**, 1813-1823.

Szaro, R.C. and Johnson, D.W. (1996) *Biodiversity in Managed Landscapes, Theory and Practice*. Oxford University Press, New York, USA. 778p.

Williams B.K., Nichols, J.D., and Royle J.A. (2002) *Analysis and Management of Animal Populations*. Academic Press, San Diego, USA.

COMPLEMENTARY WORK

Improving the performances of the occupancy parameter estimator in the case of a spatially rare species

Florent Bled and Jérôme Dupuis

Reminder: *φ=conditional occupancy parameter defined as the probability that the species is present in the quadrat ; q=probability of detecting a species in a sampled quadrat during any session (given that it is present in the quadrat) ; γ=occupancy rate ; K=number of sessions during which detection are recorded on a sampled quadrat ; J=total number of quadrats in the study area ; T=number of sampled quadrats in the study area.*

INTRODUCTION

Spatially rare species often are species that have to be closely monitored because there regularly are at least locally threatened species. Being spatially rare typically leads to low detection and consequently imprecise estimations of the occupancy parameter. The more accurate the estimators of population parameters are, the best the subsequent management decisions can be. Therefore, any way to improve estimators has to be welcome. It is now widely accepted that improving detectability can greatly ameliorate occupancy parameter estimators (MacKenzie *et al.*, 2005, 2006 ; Dupuis *et al.*, 2010) and is non-negligible parameter since the nuisance it represents when we are interested in estimating occupancy.

Several methods can be used in order to do so. In fact, it is possible to make several arrangements both before and after data collecting to ameliorate estimators' performances.

First, before data collecting, when designing the sampling protocol, several modifications can be made such as increased replication in time or in space (e.g. robust design; Pollock, 1982 ; Kendall & Nichols, 1995 ; Kendall *et al.*, 1995, 1997). Some possibilities of time replication include: multiplying the number of visits on a quadrat or dividing the global time spent on each quadrat into multiple sessions to permit a better detectability. Changes in spatial replication can include sampling a more important part of the study area. Both of these options require an augmentation of the sampling effort. Therefore, it is often needed to determine an acceptable trade-off between estimators' quality and allocated resources.

Second, once the data are collected, it is still possible to significantly improve estimators of the occupancy parameter using some available information on the parameter of interest φ

and on the nuisance parameter *q*. Indeed, in Dupuis *et al.* (2010), we have seen that under some circumstances, considering prior information on *q* can improve estimations of φ. As a matter of fact, they noticed that incorporating prior information about detectability permitted to improve estimations of occupancy rate γ when target species are highly detectable and not or few detected. However, since γ is not a parameter of the model, here we decided not to use it and to work with the occupancy parameter φ which is linked to γ as follows $E[\gamma \mid \varphi] = \varphi/[1-(1-\varphi)^J]$. Commonly, prior information is included in the form of more or less informative prior distributions on each parameter, and it should be determined how incorporating these will affect estimators. In Dupuis *et al.* (2010), remarks have been done on some estimation, but here, we directly worked on estimators. Several studies have already been made in order to determine 'optimal' sampling designs (e.g. Tyre *et al.*, 2003 ; Field *et al.*, 2005 ; MacKenzie & Royle, 2005 ; MacKenzie *et al.*, 2005 ; Pacifi *et al.*, 2009) but none of them examined estimations' improvements through the incorporation of prior information. Moreover, assessing improvements of $\hat{\varphi}$ via mean square error (as done in this paper) is new, since previous works used either bias or variance to assess the performances of the occupancy probability estimator.

As a complementary work of the conditional approach shown in this chapter, we have conducted simulation studies to determine how estimator of occupancy parameters (i.e. $\hat{\varphi}$) could be improved when the previous arrangements are made both before and after collecting data. We can expect that estimators' quality will be improved as the number of visits in each quadrat increases (temporal replication) and as a more important part of the study area is sampled. Moreover, we can expect that incorporating prior information on detection probability *q* should improve the quality of the corresponding estimator $\hat{q}$, and indirectly the estimator $\hat{\varphi}$ of the occupancy probability φ.

METHODS

Estimators' quality measurement

Mean Square Error – In order to assess estimator quality, we have chosen to use the Root Mean Square Error (R.M.S.E.). We have $RMSE(\hat{\varphi}) = \sqrt{Var(\hat{\varphi}) + [Bias(\hat{\varphi})]^2}$. This measure combines bias (difference between expected value of the estimator and the parameter) and precision in one overall measure of how "close" an estimator is to the corresponding parameter since $MSE(\hat{\varphi}) = E[(\hat{\varphi} - \varphi)^2]$.

Simulated dataset – Since it is not possible to directly calculate the bias and variance of $\hat{\varphi}$, we have implemented classical Monte Carlo methods (thus simulating datasets under the conditional model with known values of φ and q). Computation of the R.M.S.E. for $\hat{\varphi}$ and $\hat{q}$ has been based on 1,000 simulated datasets for which true values of φ and q were set (respectively denoted by φ^* and q^*). This permitted to calculate the corresponding R.M.S.E. under each condition. We have examined how estimators' quality is impacted by: i) variations on K (number of visits on each quadrat), ii) on r (ratio of the number T of sampled quadrats over the total number J of quadrats in the region) and iii) on force of prior information.

Sampling design adjustments

Time replication – One way of potentially improves $\hat{\varphi}$, is to improve $\hat{q}$ by visiting each quadrat repeatedly. It corresponds to globally increase K keeping each individual observing session the same time. Here, the underlying idea is that every new visit will permit to simultaneously obtain a more accurate estimation of all parameters (i.e. occupancy and detection probabilities). We will assess how R.M.S.E. for $\hat{\varphi}$ and $\hat{q}$ varies when K varies between 4 and 12. This will be compared for four conditions: i) when J=100, T=100, φ^* =0.1, q^* =0.1 ; when J=20, T=20, q^*=0.1 with ii) φ^* =0.6, iii) φ^* =0.45 and finally with iv) φ^* =0.3.

It is also possible to modify the number of visits K while keeping the global observation time constant. Here, the total amount of time of the visit stays the same regardless of the value of K. K varies between 1 and 8. We will only present results when J=20, T=20, φ^*=0.3 and detection probability over the whole period of visit is constant and equal to 0.7. Here, we will only provide R.M.S.E. for $\hat{\varphi}$.

Spatial replication – The other way of improving detectability estimator is to modify the ratio of quadrats sampled over the whole studied area ($r=T/J$). We have examined how R.M.S.E. react when r varies from 20 to 100% with K=10, for φ^*=0.3 and q^*=0.1. This will be done for J=100 and J=20.

It should be noted that, here, in both cases of spatial and time replication, we have used non informative priors on φ and q (that is uniform distribution on [0,1]).

Incorporation of prior information

Force of the prior – Commonly, prior information is provided as a prior illustrated by a mean and a $CI_{95\%}$ on one of the parameters. This $CI_{95\%}$ can be more or less large, indicating how strong the prior is. As seen in chapter IV, it is possible to define a prior Beta distribution that correspond to this prior and that will be used as the prior distribution on the corresponding parameter. Now, let's note ρ the 'force' of a prior. For a beta distribution, ρ is the sum of the two distribution parameters, and illustrate how strong a prior is. The higher ρ is, the stronger the prior is, and the more informative the corresponding beta distribution will be. As a matter of fact, a simple computation shows that the variance of a beta distribution with a mean μ and a force ρ is $\frac{\mu(1-\mu)}{1+\rho}$, therefore we see that the higher ρ is, the lower the variance will be (which highlights the interpretation of ρ as the force of the prior). Table 1 indicates correspondences between ρ and the prior $CI_{95\%}$ that could be provided by an expert.

Table 1. Prior $CI_{95\%}$ with a prior mean equal to 0.3 depending on 'force' ρ.

ρ	5	10	20	50	100
CI 95%	[0;0.67]	[0.03;0.57]	[0.1;0.5]	[0.17;0.43]	[0.21;0.39]

Prior on q – The most common situation is when some information on detectability is available. We have assessed how a more or less strong prior can improve estimators' quality both for φ and q. More specifically, we have considered a case where K=4, J=20, r=100%, φ^* =0.1 and q^* =0.3 (the prior on φ being non informative). The prior mean on q was set to 0.3, consequently we have used an unbiased prior since $E[q]=q^*=0.3$. Its force was increased from 2 (i.e. non informative prior) to 50. In this study, we have limited ourselves to unbiased prior for q. Therefore observed improvements concerning $\hat{\varphi}$ are in a way optimal, since it is expected that using biased priors on q will somehow deteriorate these performances, as the larger the bias is (as already observed by Dupuis & Joachim, 2006).

Prior on φ – Prior on φ might also be available but is often more likely to be biased than prior on q. We have estimated the consequences of the force of the prior information on occupancy probability both with biased and unbiased prior mean on φ, in the case where φ^*=0.2 and q^*=0.3, with J=T=20 and K=4; and with a non informative prior on q. Here, the prior bias, that is $E[\varphi]-\varphi^*$, is equal to 0.2-0.3=-0.1.

RESULTS

Sampling design adjustments

Time replication – • When the region R is fully sampled (r=100%), it does not matter if the number of quadrats is large (J=100) or small (J=20), nor what the real q and φ are (from 0.6 to 0.1 and from 0.3 to 0.1 respectively), two stages are observed. First, R.M.S.E. decreases, indicating an improvement in estimators when only few visits are made (from 4 to 8). While estimators are pretty inaccurate for low values of K, it can be greatly improved up to 36% for $\hat{\varphi}$ and 45% for $\hat{q}$. Then a second phase is quickly reached, a plateau for a number of visits higher than 8. (fig. 1, and fig. 2)

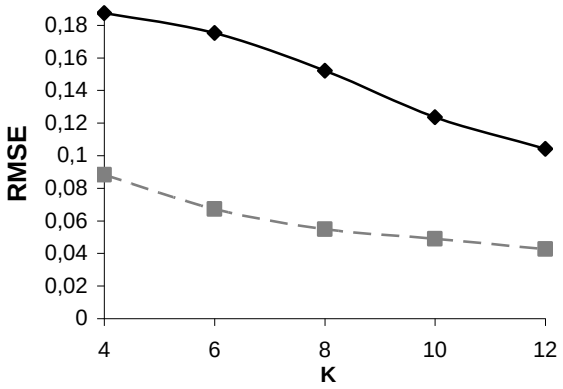


Figure 1. Root Mean Square Error for $\hat{\varphi}$ (black line) and $\hat{q}$ (grey dashed line) as a function of the number of visits K, when J=100, T=100, φ^* =0.1 and q^* =0.1.

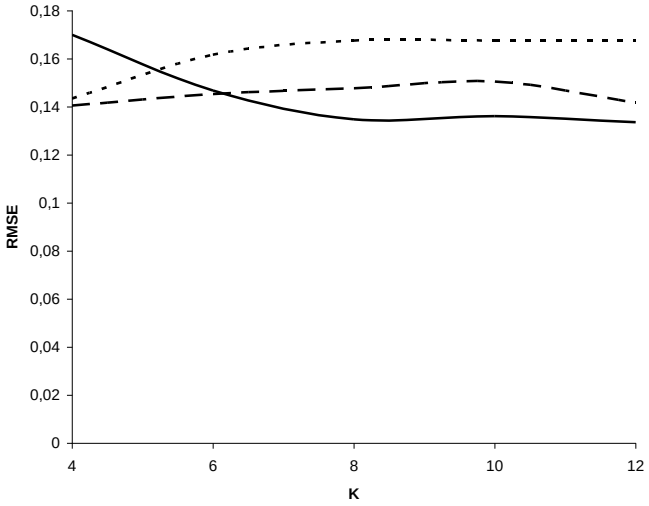


Figure 2. Root Mean Square Error for $\hat{\varphi}$ as a function of the number of visits K, when J=20, T=20, q^*=0.1 for (a), φ^*=0.6 (continuous line), (b) φ^*=0.45 (dashed line) and (c) φ^*=0.3 (dotted line).

• When we modify the number of visits *K* keeping the global observation time spent on each location constant, the expected improvement for occupancy probability estimator does not seem to be substantial (hardly 8%-decrease for RMSE between *K*=2 and *K*=8) (fig. 3). For example, this would mean that doing 4 sessions during 5 min or 8 sessions of 2min30 would not significantly change estimators' quality compared to doing 2 sessions of 10 minutes.

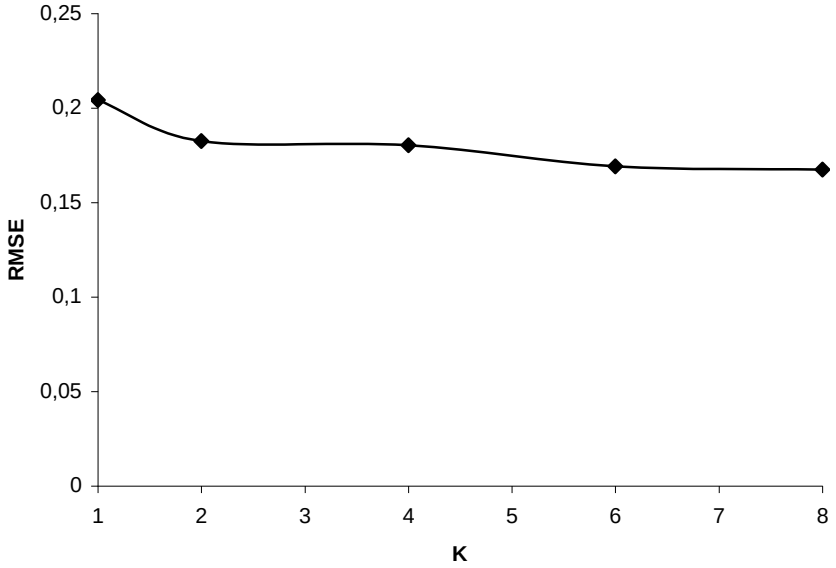


Figure 3. Root Mean Square Error for $\hat{\varphi}$ as a function of the number of visit *K*, when *J*=20, *T*=20, φ*=0.3 and detection probability over the whole period of visit is constant and equal to 0.7.

Spatial replication – When real φ and *q* vary from 0.6 to 0.1 and from 0.3 to 0.1 respectively, we observe a diminution of R.M.S.E. for φ and *q* estimators until reaching a plateau when 60-80% of the region sampled (fig. 4a). However, for a region with a small number of quadrats (*J*=20) and a low real *q* (equal to 0.1), if real φ is lower than 0.3, $\hat{q}$ is improved by an increase of the ratio of sampled quadrats while φ stays the same (fig. 4b). $\hat{q}$ is improved up to 75% when the ratio of sampled quadrats varies between 20 and 100%. In some cases (e.g. *J*=20, *K*=10, φ* = 0.6 and *q** =0.1), simply increasing this ratio can improve $\hat{q}$ by 51%. $\hat{\varphi}$ accuracy is approximately improved by 25% and can be improved up to 50% , when portion of sampled area is increasing from 20 to 100%. When *J*=100, *K*=10, φ* = 0.6 and *q**=0.3, an increase in *r* form 20 to 40% improves $\hat{\varphi}$ accuracy by 22%.

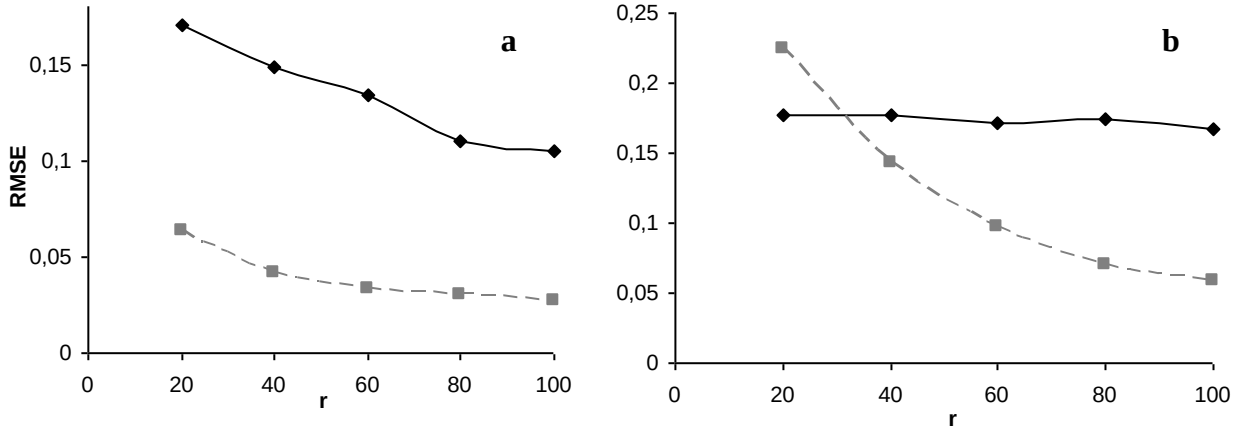


Figure 4 a, b. Root Mean Square Error for $\hat{\varphi}$ (black line) and $\hat{q}$ (grey dashed line) as a function of the ratio *r* of quadrats visited, for φ*=0.3 and q*=0.1, *K*=10 when (a) *J*=100 and (b) *J*=20.

Incorporation of prior information

Prior on q – When the prior on *q* is non informative, the R.M.S.E. of $\hat{q}$ is quite high, with a value of 0.15 (fig. 5). A moderate prior (ρ=10, i.e. $CI_{95\%}$=[0.03;0.57]) is enough to improve $\hat{q}$ by 49% relatively to a non informative prior (i.e. ρ=2), while a strong prior (ρ=50, i.e. $CI_{95\%}$= [0.17;0.43]) improves it by 87%.

With ρ=10, the improvement of occupancy probability estimator is equal to 28% relatively to a non informative prior, and R.M.S.E. for $\hat{\varphi}$ even drops by 50% for ρ=50. In a non informative situation, R.M.S.E. is approximatively 0.18. Interestingly, with a decrease by 27% (R.M.S.E.=0.13), 40% (R.M.S.E.=0.11) and 50% (R.M.S.E.=0.09) estimator's performances rapidly increases as the force of the unbiased prior augments (respectively ρ=10, ρ=20 and ρ=50), highlighting the interest of incorporating such information when available.

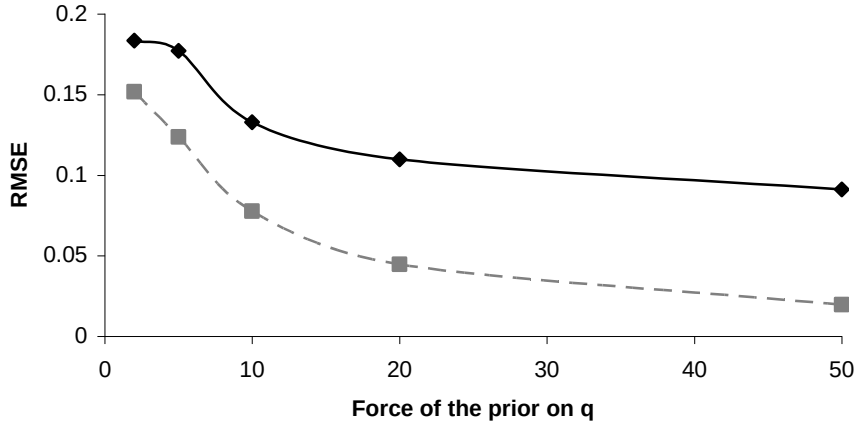


Figure 5. Root Mean Square Error for $\hat{\varphi}$ (black line) and $\hat{q}$ (grey dashed line) as a function of prior 'force' ρ, when *J*=20, *T*=20, *K*=4, φ*=0.1 and q*=0.3.

Prior on φ – As expected, prior information on φ permits to greatly decrease RMSE of $\hat{\varphi}$ (up to 88% for ρ=50 for a non-biased prior relatively to a non informative prior, and up to 52% for ρ=20 for a biased prior) (fig. 6). It is interesting to note that this is true both for unbiased and biased (prior's mean on φ is 0.3 instead of 0.2) priors. While an increase in the force of the prior constantly improves $\hat{\varphi}$ in the case of a non-biased prior, for really strong biased priors (ρ>50) improvement of the estimator will start to be less efficient but $\hat{\varphi}$ stays really good (around 50%).

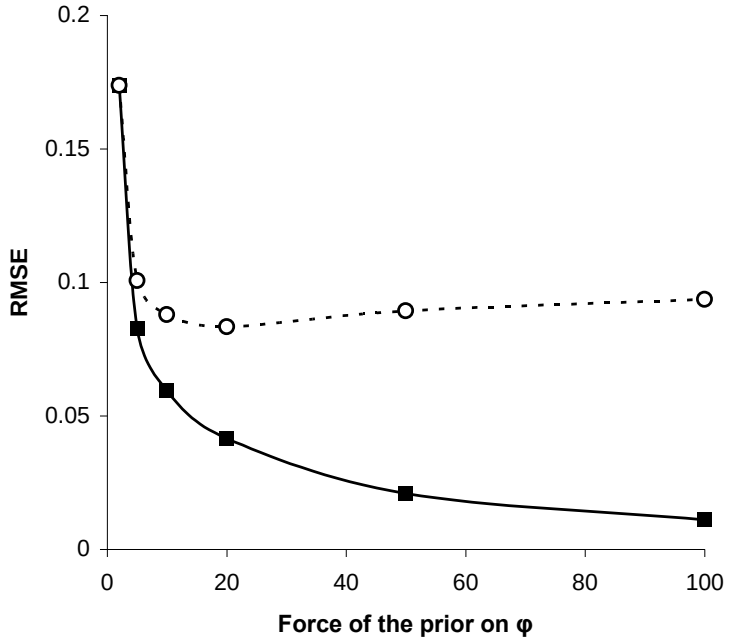


Figure 6. Root Mean Square Error for $\hat{\varphi}$ using a non biased prior (black line, mean prior equals 0.2) and a biased prior on φ (dotted line, mean prior equals 0.3) as a function of prior 'force' ρ, when J=20, T=20, K=4, φ*=0.2 and q*=0.3.

DISCUSSION

Sampling design adjustments

As seen here, some few modifications to classic study protocols can permit to greatly improve estimators' performances. Time replication can be pretty efficient, and pretty logically sampling a more important part of the study area rapidly allows an amelioration of estimators. These modifications do not have to be highly resource consuming to be efficient. As a matter of fact, while in some case, investing too much effort in data collecting (e.g. increasing K over 8 when φ*=0.3, J=T=20 and q*=0.1), if done unwisely, can lead to a deterioration of estimators' performances. Several protocols have been conceived in order to

estimate occurrence, abundance or occupancy rate accounting for detectability issues, but the most appropriate study can be different depending on the true value of φ and q. A poorly designed study may yield to no useful information and may only succeed in wasting resources. Therefore preliminary studies can be really useful in order to adopt the best protocol for a specific question and for a given species.

Incorporation of prior information

The use of prior information can also improve estimators, even if this prior is not strongly informative. While some priors we have used here might seem rather strong, they are not unusual to be obtained by an expert. For example, we will see in the next chapter several cases of the use of priors with a force equivalent to 50. Prior on q permits to improve $\hat{q}$, which seems logical, but more interestingly it also ameliorates the performances of $\hat{\varphi}$. While a really moderate prior on q (ρ=5) does not really improve $\hat{\varphi}$; increasing the force of an unbiased prior on q can rapidly improve $\hat{\varphi}$. Therefore, with this model, prior on q rapidly permits to obtain a better estimator of φ. We have also seen that unbiased informative priors on φ can permit to improve estimators, and surprisingly even a biased prior can significantly ameliorate the assessment of population parameters. Furthermore, it would be interesting to extend this work to see how biased a prior can be and still improve the estimators (work in progress). Increasing prior's bias with a set prior force would permit to evaluate this. We could expect that estimators' performances will decrease once the force of the biased prior will be greater than the weight of the data, and also as the bias increases.

In their 2002's article, MacKenzie *et al.* suggested that the number of surveys required to provide a 'reasonable' estimate of occupancy based upon a simulation study to be at least 2 if occupancy was over 0.7 and detection probabilities (in a single survey) over 0.3. Tyre *et al.* (2003) also used simulation results in order to give some advice on the number of surveys required at each sampling unit. They concluded that it is better to increase ratio of sampled quadrats rather than K if the detection probability is high. However, when detection probability decreases, they suggested that more surveys per unit should be conducted. In 2005, Field *et al.* have investigated how study design permits to improve estimators' quality based on the power of a study to detect a decline in the level of occupancy over a 3-year period. Using simulation methods, they concluded that K=2 or 3 would usually be sufficient unless occupancy was high or detection probability was low. Using the unconditional model

presented in MacKenzie *et al.* (2002), MacKenzie & Royle (2005) reached the conclusion that attempting to survey as many sites as possible may not be the most efficient use of resources, and that surveying fewer sites more often may result in a more precise estimate of occupancy. In the end, they recommend that K=3 should be considered as a minimum value when detection probability is over 0.5, and be greater when it is smaller. Here, we used a conditional approach, meaning that our parameters has a slightly different signification then the ones they employed. Therefore, we can not directly compare those values, but we agree that increasing K or r indefinitely is a waste of resources and sampling design had to be carefully studied before final sampling. Each specific condition calls for a specific adapted design. Moreover, when available, prior information (on occupancy probability, detectability or even presence of the species in the area of study even if not detected) should be taken into account. It is quite straightforward to do so in the Bayesian framework. Nevertheless, surprisingly, only few examples of actual applied studies can be found in the literature and it could be interesting to raise populations' managers' awareness to these kinds of statistical procedures.

A further discussion of these results appears in the last chapter of this thesis but we can already highlight that few modifications sampling design, and the use of some prior information can greatly improved estimators' performances under specific condition. This has to be preliminarily studied in order to maximize results' quality and therefore the subsequent management decisions that would be taken.

REFERENCES

Dupuis J., Bled F. & Joachim J., 2010. Estimating the occupancy rate of spatially rare or hard to detect species: a conditional approach. *Biometrics*. (*in press*)

Dupuis J. & Joachim J., 2006. Bayesian estimation of species richness from quadrat sampling data in the presence of priori information. *Biometrics*. **62**: 706–712.

Field S.A., Tyre A.J. & Possingham H.P., 2005. Optimizing allocation of monitoring effort under economic and observational constraints. *Journal of Wildlife Management*. **69**: 473–482.

Kendall W.L. & Nichols J.D., 1995. On the use of secondary capture-recapture samples to estimate temporary emigration and breeding proportions. *Journal of Applied Statistics*. **22**: 751–762.

Kendall W.L., Nichols J.D. & Hines J.E., 1997. Estimating temporary emigration using capture-recapture data with Pollock's robust design. *Ecology*. **78**: 563–578.

Kendall W.L., Pollock K.H. & Brownie C., 1995. A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics*. **51**: 293–308.

MacKenzie D.I. & Royle J.A., 2005. Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*. **42**: 1105–1114.

MacKenzie D.I., Nichols J.D., Lachman G.B., Droege S., Royle J.A. & Langtimm C.A., 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*. **83**: 2248–2255.

MacKenzie D.I., Nichols J.D., Royle J.A., Pollock K.H., Bailey L.L. & Hines J.E., 2006. *Occupancy Estimation and Modeling*. Academic Press. San Diego, California, USA. 324p.

MacKenzie D.I., Nichols J.D., Sutton N., Kawanishi K. & Bailey L.L., 2005. Improving inferences in population studies of rare species that are detected imperfectly. *Ecology*. **86**: 1101–1013.

Pacifici K., Conroy M., Cooper R. & Peterson J., 2009. Occupancy estimation within an adaptive sampling design: A calibrated Bayesian approach to evaluating an integrated estimator for rare or elusive species. *EURING 2009 Analytical Meeting & Workshop*. 14-20 September, 2009. Pescara, Italy.

Pollock K.H., 1982. A capture–recapture design robust to unequal probability of capture. *Journal of Wildlife Management*. **46**: 757–760.

Tyre A.J., Tenhumberg B., Field S.A., Niejalke D., Parris K. & Possingham H.P., 2003. Improving precision and reducing bias in biological surveys by estimating false negative error rates in presence–absence data. *Ecological Applications*. **13**: 1790–1801.

CHAPTER V

1

IMPACT OF CLIMATIC VARIATIONS ON BIRD SPECIES OCCUPANCY RATE IN A SOUTHERN EUROPEAN FOREST

BLED Florent[1,*], JOACHIM Jean[2], DUPUIS Jérôme[3]

[1] Laboratoire Evolution et Diversité Biologique, Université Paul Sabatier, France

[2] INRA-CEFS, BP 27, 31326, Castanet-Tolosan, Toulouse, France

[3] Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse, France

(Biodiversity and conservation, in press)

* Corresponding author (e-mail: bled@cict.fr)

Abstract

Species that are affected by climatic variations can undergo modification in range and/or abundance. Knowing how individuals or species occupy their habitat is essential to understand how species use their environment, and detecting variations that might affect this use can be determinant in species management.

Hierarchical modeling is regularly used to assess for occupancy rate (i.e. proportion of patches occupied in a region), particularly when it is required to consider detectability-related issues. The present study is the first application of the conditional model presented in Dupuis *et al.* (2010), which is applied in the case of a heterogeneous area that might be divided into homogeneous sub-areas. Their approach is used to study the impact of three consecutive particularly cold winters on a selected set of bird species in a forest of southern France in the context of available prior information on birds detectability.

We examined a limited range of factors that might influence the response of some bird species to climatic. We considered the case of sedentary, partially migrating and migrating species. We also assessed if the biogeographical origins of the different species affect their occupancy rates. Globally, changes in occupancy rates between 1985 and 1987 indicates for the first time a continentalization of the regional forest fauna, reflected by the expansion of Palearctic and Turkestano-european faunistic type species, with depletion or extinction of European, Turkestano-mediterranean and Mediterranean sedentary species. We have also shown the importance of prior information.

Keywords: Climatic changes, detectability, hierarchical modeling, occupancy rate, prior information.

Introduction

Two of direct effects of climate change that species will likely have to face are an increase in mean temperature) (e.g. Malcom *et al.*, 2006) and an increase in the frequency of extreme temperatures (e.g. Colombo *et al*., 1999). Species' responses to these climatic changes may include evolution or adaptation, modification in range and abundance, or extinction (Ackerly, 2003 ; Parmesan, 2006).

Most of the examples of rapid adaptation to climate change lie in evolution toward higher frequencies of already existing heat-tolerant genotypes in the interiors of species' ranges (Parmesan, 2006). While adaptation can occur after a significant amount of time, for punctual variation of the climate, migration and extinction are more likely (Huntley, 1991 ; Davis & Zabinski, 1992 ; Coope, 1994). Extinction or increased risk of extinction due to climatic changes have been documented in a wide range of taxa such as amphibians (Pounds *et al.*, 1999, 2005), tropical corals (Hoegh-Guldberg 1999; Wilkinson 2000), mammals (e.g. Arctic polar bears, Derocher ; 2005, Stirling *et al.* 1999) and birds (e.g. Antarctic Adélie Penguins, Ainley *et al.*, 2003 ; Croxall *et al.*, 2002). In fact, if an unusual variation in temperature occurs, such as a significantly colder winter, species that are not adapted to cold temperature and that have not migrated might have to undergo extreme conditions and potentially go locally extinct. Range shifts are expected to occur poleward and upward. Several species of Lepidoptera (butterflies and moths) have undergone an expansion of northern boundaries situated in Finland (Mikkola, 1997), Great Britain (Hill *et al.*, 2002), and across Europe (Parmesan *et al.*, 1999). Some bird species have been shown to have a northern shift of their range boundaries (Thomas & Lennon, 1999). Some species have even shown elevational shifts in response to climate changes (e.g. lowland birds, Pounds *et al.*, 1999, 2005 ; alpine flora, Grabherr *et al.*, 1994, Pauli *et al.*, 1996).

In this context, understanding how individuals or species occupy their habitat and use their environment (Krebs, 1978) will be essential to predict population response to extreme climatic conditions and take suitable management dispositions (Kendall, 2001; Williams *et al*., 2002 ; Nichols, 2004 ; MacKenzie *et al.*, 2006). Site occupancy modeling has been thoroughly developed during the last decade, leading to powerful methods to estimate site occupancy, which in turn allows estimation of population parameters such as colonization and extinction probabilities. These models have proven to be reliable for common ecological data (i.e. detection-non detection), and when the area of interest is large (such as a country) or relatively small (e.g. a forest), or when the considered area is divided in an important number of quadrat (i.e. sub-unit of the global area that might be sampled) or not.

Site occupancy rate is one of the key variables for determining how species use their environment. While density and abundance carry information on the population state, occupancy rate describes how species use the spatial environment they live in. It is defined as the proportion of patches occupied and is a state variable in various metapopulations models (e.g., Levins, 1969, 1970 ; Lande, 1988 ; Hanski, 1997). Site occupancy rate by definition has to be considered in a 'closed environment', i.e. with a finite number of locations (termed sites or quadrats). While this has been well studied when there is a large number of sites, improvements are still needed when the number of sites is small (MacKenzie *et al*., 2006).

Estimating occupancy rate using hierarchical modeling allows accommodation of issues related to detectability. We now can account for the fact that when a species is not detected it could correspond to an effective absence or just a failure in detection. Modifications of the original approach of MacKenzie *et al.* (2002) allowed for a more accurate estimation of occupancy rate in the specific case of a small homogeneous area with a small total number of quadrats (e.g. MacKenzie *et al.*, 2006) or when species are particularly hard to detect (Dupuis *et al.*, 2010). This difficulty in detection can emerge from individual's characteristics (e.g.

cryptic abilities) or simply because a species has a low presence probability, which is typical of endangered species.

Here, we apply the conditional model presented in Dupuis *et al.* (2010) to the case of a heterogeneous area that can be divided into homogeneous sub-areas (e.g. edge and inner part of a forest). One of the main characteristics of this model is that it uses a conditional approach for estimating the quantity of interest (here, occupancy rate); conditioning bearing on the presence of target species in the region of study. The unconditional approach (MacKenzie *et al.*, 2006) does not presuppose that the target species is present in the region of interest. Conditional approach has been proven to be a necessary approach in the case of undetected species which are known to be present in the area from another source of information (such as net capture) (Dupuis *et al.*, 2010). Moreover, through the use of a Bayesian approach, we can take advantage of that some prior information on detectability is available, since it is expected that this significantly improves the precision of estimates when the target species has been very few or not detected, and is known to be easily detectable (Dupuis *et al.*, 2010). The data used in this study present several such examples. Furthermore, in each instance where a species is not detected in our dataset it is known from external sources that it was in fact present in the study area.

We focus on several factors that might influence the response of bird species to climatic changes (here the succession of three significantly colder winters). First, occupancy rate in the inner forest might be less affected by the decrease in temperature than in the edge of the forest because more protection is available for individuals. Evolution of occupancy rates might be crucially different between the external and the inner parts of a forest. For example, species can have preferences into one or the other part, and if no distinction is made between these regions, one could not see a modification of occupancy. Second, we compare sedentary, partially migrating and migrating species. We expect migratory species occupancy rate to be

less impacted by cold winters than that of sedentary species, because they are not directly affected by local low temperatures at reproduction sites. Specifically, we predict sedentary species' occupancy rate to decrease. Third, we compare species with different bioclimatic affiliations which illustrate tolerance of a species to cold. We predict that species with a more continental bioclimatic affiliation will not experience a dramatic decrease in occupancy rate compared to other species. Finally; in the discussion, each of the conclusions resulting from this local application will be examined in the light of what happened at the continental scale.

Material & Methods

Location – We focus on data collected in May 1985 and May 1987 in order to estimate occupancy rates of nesting bird species present in the Montech forest (located near Toulouse, France), and the impact of 3 successive particularly cold winters. Montech forest has an area of 900ha and is part of the French National Forest System (*Office National des Forêts*). While this forest has a mixed management regime of mature and coppiced forest, it is now only managed as a mature forest. This forest is dominated by oak species (*Quercus robur*, *Quercus sessiliflora* and *Quercus pubescens*). The undergrowth is composed of hawthorns (*Crataegus)*, blackthorns *(Prunus spinosa)*, brooms *(Fabaceae)*, and heaths *(Erica)*. Some Coast Douglas-firs *(Pseudotsuga menziesii)* and Maritime Pines *(Pinus pinaster)* plots are present in the south part of this forest. The study area is proximate to continental, Mediterranean, alpine and Atlantic ecoprovince (Fig.1).

Data are a part of a broad study of bird populations in the Midi-Pyrénées region of southwest France. Data collection began in 1985 (Joachim *et al.*, 1997). Montech forest underwent 3 colder-than-usual winters in 1985, 1986 and 1987. During these winters, mean

monthly temperatures dropped up to 5°C below the average monthly temperature of 14 previous years, and 3°C below the coldest temperatures during this reference period (Fig. 2). These three winters correspond to a hinge period. In contrast, from 1965 to 1985, winters were fairly mild. Empirical observations suggest that these mild temperatures led to a colonization of Mediterranean and a retreat of continental bird species. The exceptional 1985-1987 period contrasted strongly and that likely impacted the fauna potentially leading to a strong re-continentalization. While species might easily overcome one colder than usual winter, the succession of 3 cold winters on end can be expected to have consequences for bird species.

Data and species list – We focus on to the estimation of occupancy rate and detection probabilities of 10 bird species of the overall 47 detected species. The list of the species of interest is denoted by L . Here, the list of chosen bird species, liable to be present in the forest of Montech, in spring 1985 and 1987 includes 10 species: whitethroat, chiffchaff, nightingale (which are migratory species), Eurasian nuthatch, green wood pecker, short-toed treecreeper (which are sedentary species), and finally song thrush, common blackbird, European robin, mistle thrush (which are partial migrants). Partial migrant species are species with overlapping reproductive and wintering areas. Migratory status and biogeographical origins are presented in table 2. Moreover, for each species, prior information on detection probability is available (Table 3).

Montech forest, with a surface area of 900 hectares has been divided into A homogeneous subregions of biological interest, here A=2 with the two sub-regions being the inner forest (a=1) and the edge forest (a=2). A sampling is performed in each sub-region R_a. The sub-region R_a is divided into J_a spatial units, here termed quadrats. They are also called sites in

the literature. A subset of T_a quadrats is sampled. Draw is usually performed at random so as to have a sample representative of the whole region R_a. We assume that these quadrats are of equal area. Twenty-two quadrats are located in the inner forest (J_1=22), while 18 quadrats correspond to forest's edges (J_2=18). In 1985, all quadrats were sampled (T_1=J_1=22 ; T_2=J_2=18), while in 1987 only 14 quadrats were sampled in inner forest (T_1=14) and 6 in edge forest (T_2=6).

Information about the presence of nesting species was provided by acoustic recognition of singing males according to the following procedure. Listener spent a twenty minutes at each station, listening for birds. This time spent on each quadrat was divided into *K* equal sessions, here *K*=4, during which detection are recorded. The number of quadrats in region R_a in which species *s* has been detected is denoted by $V_s(a)$. Total number of sessions in region R_a during which species *s* has been detected is denoted by $W_s(a)$. These data are given in Table 1. All records were made following the listening point method (EPS, Spitz 1974, EFP, Blondel, 1975) over a 500-meters sided grid. Length of the side of a quadrat (that is 500 meters) has been chosen so that the listening range (a disk of 150-meter radius) is strictly included within the quadrat. Consequently, from a station performed in a given quadrat, the observer cannot hear songs of (males) conspecifics that would nest in adjacent quadrats; consequently, if the song of a male of a given species *s* is detected in the quadrat *j*, thus the species *s* is present in this quadrat. Further details of the protocol can be found in Decamps *et al.* (1987).

The dataset of observations is afterwards denoted by $\boldsymbol{y}$; therefore $\boldsymbol{y} = \{\boldsymbol{y}_s(a); s \in L\ , a = 1,..., A\}$, where $\boldsymbol{y}_s(a)$ denotes the observation related to species *s* in region R_a. A possible record of $\boldsymbol{y}_s(a)$ for a region R_a could be $\boldsymbol{y}_s(a)$=4021003. Here, $V_s(a)$=4, $W_s(a)$=10. Finally, the number of detection of species in the sampled quadrat *j* is denoted by y_{sj}.

Modeling

Underlying processes and missing data structure – To specify the missing data structure inherent in quadrat sampling data, we view the record $\boldsymbol{y}_s$ as the result of two processes: one is related to the presence-absence process (in quadrats), and the other is related to the detection process. Such a formalism also allows us to formulate rigorously the biological assumptions made and to introduce, in a natural way, the parameters of biological interest.

We denote by z_{sj} the indicator of presence of species s in quadrat j; therefore, $z_{sj} = 1$ if species s is present in quadrat j, and 0 otherwise.

Conditioning on z_{sj}=1, x_{sj} is the number of times that species s has been detected in quadrat j during the K sessions. It is clear that $x_{sj} = 0$ and $y_{sj} = 0$ do not have the same meaning. The event $x_{sj} = 0$ means that the species s (present in quadrat j) has not been detected in it; while the event $y_{sj} = 0$ resorts two distinct situations. Missing can occur in different circumstances. First, when a species s has not been detected in quadrat j (that is when $y_{sj} = 0$), it is clear that z_{sj} is missing; the event $y_{sj} = 0$ covers in fact two exclusive situations: either species s is present in quadrat j but has not been detected, or it is not present in quadrat j (and cannot have been detected). Formally, one has the equivalence $(y_{sj} = 0) \Leftrightarrow (z_{sj} = 1 \text{ and } x_{sj} = 0) \text{ or } (z_{sj} = 0)$. Conversely, when $1 \le k \le K$, one has $(y_{sj} = k) \Leftrightarrow (z_{sj} = 1 \text{ and } x_{sj} = k)$. Second, z_{sj} is also missing, when quadrat j is not a part of the sampled quadrats.

Notation – Throughout the paper $p(.)$ denotes a probability mass function. The local occupancy rate of species s in region R_a is denoted by γ_{sa}; therefore, one has:

$$\gamma_{sa} = \frac{1}{J_a} \sum_{j \in R_a} z_{sj}$$

We apply the conditional model of Dupuis *et al.* (2010) to each sub-region R_a for estimating γ_{sa}. Biological assumptions related to the occurence of species in the J_a quadrats are supported by the values of z_{sj}, and those related to detection in the T_a sampled quadrats by the values of x_{sj}.

Modeling detectability – We assume that the probability of detecting species s in quadrat j does not depend on its detection in the other quadrats. Considering the experimental protocol, size and distance between sampled quadrats , this is a realistic assumption.

For any quadrat j of R_a, we assume that conditionally on z_{sj}=1, x_{sj} follows a Binomial (K, q_{sa}). Here q_{sa} represents the probability of detecting species s in quadrat j (located in region R_a) during any session (given that it is present in quadrat j). Detection probability can vary strongly among species, and can also differ depending on the subarea considered. We also can define the parameter $\mu_{sa} = \Pr(x_{sj} \geq 1 \mid z_{sj} = 1) = 1-(1-q_{sa})^K$ which represents the probability of detecting species s in quadrat j (conditionally on its presence) over the K sessions of a visit. This parameter is commonly the parameter on which prior information is available.

Modeling occupancy – Let s be a species present in the study area R_a, we set $\boldsymbol{z}_s(a) = \{z_{sj}; j \in R_a\}$. Since we adopt a conditional approach, $\boldsymbol{z}_s(a)$ is distinct from the null vector. The p.m.f. (probability mass function) of $\boldsymbol{z}_s(a)$ is:

$$p(\boldsymbol{z}_s(a) \mid \varphi_{sa}) = \frac{\varphi_{sa}^{N_s(a)}(1-\varphi_{sa})^{J_a - N_s(a)}}{1-(1-\varphi_{sa})^{J_a}} \qquad (3.3)$$

where $N_s(a)$ is the number of quadrats (of R_a) where species s is present. φ_{sa} is the probability that species s is present in quadrat j of R_a, given that it is present in at least one other quadrat of R_a. Note that φ_{sa} is a conditional occurrence parameter (contrary to what appear in MacKenzie *et al.*, 2006). Also γ_{sa} is not a parameter of the model, in contrast to φ_{sa}; but they are linked, since $E[\gamma_{sa} \mid \varphi_{sa}] = \varphi_{sa}/[1-(1-\varphi_{sa})^{J_a}]$. Due to the conditioning, the random variables z_{sj} are not independent. However, a certain form of conditional independence between the z_{sj}'s holds. Let i, j be any two quadrats of R_a, z_{si} and z_{sj} are independent, if species s is present in at least one other quadrat of R_a (distinct from i and j).

Therefore, from a biological point of view, we assume that the presence of species s in quadrat i is not affected by the presence of that species in quadrat j (given that it is present elsewhere in R_a); this assumption is standard (disregarding the conditioning).

Prior information on species detectability – These data are part of a large study in the Midi-Pyrénées (i.e. South-West of France), therefore we have access to prior information on the probability of detecting species during the entire visit. Information on this detectability is also available from the literature. It should be noted that, astonishingly, in the literature Bayesian studies using informative priors are not as common as it could be expected. Here, since this information is available and can bring substantial improvements to estimates, we have logically used it in the process of occupancy rate assessing.

Typically, litterature provides information on species detectability and these species can be defined for example as highly detectable or hardly detectable (Blotzheim & Bauer, 1985, 1988, 1993, 1997). Dupuis & Joachim (2006) have defined categories of detectability in order to classify bird species depending on this information for 1985. Using information provided

by external data, they have found the Beta prior distribution for detection probability μ_{sa} corresponding to each of these categories.

Here we use 4 categories: low, intermediate low, intermediate high and high detectability. The mean and $95\%_{CI}$ of the prior distributions for each category, respectively, are 0.2 [0 – 0.4], 0.4 [0 – 0.65], 0.6 [0.3 – 0.8] and 0.8 [0.65 – 1]. These specific priors for each species are presented in Table 3. We have used the prior means and $CI_{95\%}$ intervals on μ_{sa} to determine coefficients of the corresponding Beta distribution by the use of a binary search algorithm (see MatLab code in appendix 1). Then, these coefficients are used to estimate coefficients of the prior Beta distribution on q_{sa} (the probability that a species is detected during each session of a visit) through the use of classical Monte Carlo methods (see MatLab code in appendix 2). Finally, the prior information on q_{sa} is used to estimate γ_{sa}.

Prior information on detectability for some species varies between 1985 and 1987 and/or between areas (see Table 3). Modification of detectability classification for some species in 1987 is due to regional specificities or changes in the abundance of individuals. For example, the song thrush is less inclined to sing as the number of conspecifics in the neighborhood decreases. In the case of the whitethroat in the forest interior, no estimation has been done because it is known that this species does not live in this type of environment (therefore, no prior has been specified in Table 3).

Likelihood – The parameter of the model is $\theta = \{\varphi_{sa}, q_{sa}; s \in L, a = 1,.., A\}$. The likelihood of θ based on $\boldsymbol{y} = \{\boldsymbol{y}_s(a); s \in L, a = 1,.., A\}$ is:

$$L(\theta; \boldsymbol{y}) = \prod_{s \in L, a=1,...,A} p(\boldsymbol{y}_s(a) \mid \varphi_{sa}, q_{sa})$$

Note that we implicitly assume that species behave independently with respect to occupancy. For any species s such that $\boldsymbol{y}_s(a)$ is different from the null vector:

$$p(\boldsymbol{y}_s(a)\,|\,\varphi_{sa},q_{sa}) = \frac{\rho_{sa}\varphi_{sa}^{V_s(a)}q_{sa}^{W_s(a)}[1-q_{sa}]^{U_s(a)}[(1-q_{sa})^K\varphi_{sa}+1-\varphi_{sa}]^{T_a-V_s(a)}}{1-(1-\varphi_{sa})^{J_a}} \qquad (1.1)$$

where $U_s(a) = KV_s(a) - W_s(a)$, and $\rho_{sa} = \prod_{j\in R_a}\binom{K}{y_{sj}}$; moreover, if $\boldsymbol{y}_s(a)$ is a null vector, we have:

$$p(\boldsymbol{y}_s(a)\,|\,\varphi_{sa},q_{sa}) = \frac{((1-q_{sa})^K\varphi_{sa}+1-\varphi_{sa})^{T_a}-(1-\varphi_{sa})^{J_a}}{1-(1-\varphi_{sa})^{J_a}} \quad . \qquad (1.2)$$

Computational details are available from the authors on request. The formulae (1.1) and (1.2) allow computation of the probability (in function of J_a, T_a, K, φ_{sa} and q_{sa}) that a species s present in R_a has not been detected (or rarely detected). This occurs with relatively high probability when species s is spatially rare and/or hard to detect (Dupuis *et al.*, 2010).

Estimating the occupancy rate – We use Monte Carlo Markov Chain methods (MCMC) in order to obtain the bayesian estimations of γ_{sa}. Here, we are especialy interested in the estimation of the occupancy rate but in some cases it could be required to also estimate the parameters q_{sa} and φ_{sa}. The MCMC algorithm, taking advantage of the missing data structure, is implemented on $(\theta,\boldsymbol{z_m})$ where $\boldsymbol{z}_m = \{\boldsymbol{z}_s^{mis}(a); s\in L\ , a = 1,..,A\}$ and $\boldsymbol{z}_s^{mis}(a)$ is the set of the missing z_{sj}'s related to s (where *j* is in R_a). The bayesian estimate of γ_{sa} is easily obtained by applying the ergodic theorem (details are omitted). For further details, refer to Dupuis *et al.* (2010). The MatLab code for estimating γ_{sa}, φ_{sa} *and* q_{sa} is provided in Appendix 3. In order to estimate occupancy rates for both areas in the forest (edge and inner forest), this algorithm has been applied independantly to data from both areas, for each species, and for both years.

We also compare differences between estimates obtained under both MacKenzie *et al.*, (2006) unconditional and Dupuis *et al.* (2010) conditional approaches in the case of not

detected species that are known to be present and easily detectable. This comparison is done with non informative and informative priors on the detection probability.

Results

Globally, occupancy rates variations between 1985 and 1987 do not seem to be mainly directed by the migratory status (Table 4). For example, while some sedentary species' occupancy rates have decreased (e.g. the short-toed treecreeper), others have increased (e.g. the Eurasian nuthatch). The same result exists for migrating species and partial migrants (e.g. mistle thrush occupancy rates increased contrary to the song thrush).

Overall, occupancy rates in interior and edge forests presented the same type of changes (i.e. increase or decrease), but the extent of these variations can differ between these two sub-areas. For example, while the song thrush and the nightingale greatly retreat between 1985 and 1987 in the edge forest, this variation is less important in the inner forest. On the other side, increase in mistle thrush population was more important in the inner forest than in the edge forest, apparently revealing a protecting role of the inner forest.

In the end, variations of occupancy rates seem to be mainly driven by biogeographical origins with an occupancy rate increase of European and continental species (e.g. mistle thrush, Eurasian nuthatch, European robin) and a decrease of Mediterranean and Atlantic species (e.g. nightingale, song thrush).

We observe that the bayesian non informative estimates of the occupancy rates are precise when the number of detections is high. Moreover, in 1987 and in edge, most species have

been few detected (except the common blackbird, the chiffchaff, and the mistle trush), and, not surprisingly, we observe that most of non informative estimates are imprecise.

We also observe that taking into account the prior information available on detectability significantly improves the precision of estimates when the target species has been very few detected ($W_s(a)$=1) or has not been detected ($W_s(a)$=0): the green woodpecker, the nightingale and the whitethroat are concerned by this remark (Tables 4 and 5). In Dupuis *et al* (2010), this observation has already been mentioned, and indicated that this improvement is particularly significant when the target is highly detectable.

Finally, we observe that taking into account the prior can make decrease the precision of the bayesian estimate. For the mistle trush (in 1987 and in inner forest), the range of the 95 % posterior credible interval is 0.37 in non informative situation, and 0.46 in informative situation. This unusual phenomenon typically occurs when the two following conditions are fullfilled: the weights of the data and the prior are similar, and the (informative) prior mean of the parameter of interest and its (non informative) posterior mean are very different, which indicates a `conflict' between the prior and the data. In our case, the prior variance of q_{sa} in informative situation (namely 0.075^2) is effectively relatively close to the posterior variance of q_{sa} in non informative situation (namely 0.084^2), while the (informative) prior mean of q_{sa} (namely 0.21) is effectively very far from the non informative posterior mean of q_{sa} (namely 0.66).

Use of priors here can be essential to obtain a better estimation of occupancy rates. For example, raw data in edge forest are identical for the whitethroat (1985) and for the green woodpecker (1987). With $V_s(a)$=2 and $W_s(a)$=2 for these 4 species, we would have failed to detect variation in occupancy rate without prior information (e.g. for the whitethroat), or we could have underestimated the extent of the species occupancy rate variation (e.g. for the

green woodpecker). We can highlight the importance of including prior information on detectability when available in the case of low detection data and easily detectable species.

For a species not detected by the quadrat sampling, but for which it is known that it was in fact present in the region of study R_a (e.g. the Green Woodpecker in 1985 or the Nightingale in 1987 in edge forest), Tables 6a and 6b show that ignoring such an additional information (by using the unconditional MacKenzie *et al.*, 2006, approach) will underestimate γ_{sa}. In non informative situation, distances between both estimates are very important when the region of study has been entirely sampled (J=T=18) (Table 6a), and less important (though not negligible) in the opposite case (J=18, T=6) (Table 6b). In informative situation, distances in absolute values are definitely less important (though not negligible), regardless the region of study is totally or partially sampled. Nevertheless, we note that when J=T=18 the ratio between both estimates is particularly high (namely >3.5). To be complete, let us add that when the target species has been well detected, estimates obtained under conditional and unconditional approaches are quite similar.

Discussion

Globally, occupancy rates variations between 1985 and 1987 do not seem to be mainly directed by the migratory status and no important differences between interior forest and edge forest has been shown. The main factor responsible for species' response to climate changes appeared to be their biogeographical region of origin.

Migratory status - Contrary to what we expected, the species migratory status did not seem to be an important cause of differences in changes of occupancy rates under climate changes. Evolution of occupancy rates as a response to climate changes could have been affected

because of direct effects on bird species of these changes, or secondary consequences. For example, decrease may be due to the direct impact of rigorous winters on populations (especially for sedentary species) or indirect impacts such as a decrease in available resources (particularly for migratory species that should not be affected by the direct consequences of cold in the region). An increase of occupancy rates or a less important impact of the specific regional climatic events might be expected because other species might be negatively impacted by cold winters and therefore leave unused resources and part of ecological niche. Variation in population is the result of trades-offs between these negative and indirect positive consequences and the influence of other external factors (e.g. a population already in a decrease phase).

The impact of climatic extremes on bird densities is well-known (Wiens, 1981, Grzybowsky 1983, Balda *et al.*, 1983 ; Hejl & Beedy, 1986). According to these authors, rigorous winters affect sedentary but not migratory bird species. In our study area, Palearctic and Turkestano-european faunistic type species expanded, while European, Turkestano-mediterranean and Mediterranean sedentary species experience population reductions or extinction. This signifies a continentalization of the local avifauna. However, not all of the sedentary birds were negatively affected by cold winters. Woodpeckers, Eurasian Nuthatch and tits population increased. Fauna continentalization can also be seen with migrants. Whitethroat, Tree Pipit, Spotted Flycatcher, but also Yellow Wagtail, Red-backed Shrike and Garden Warbler have increasing occurrence frequencies.

The absence of clear impact of the migratory status can be due to the specific characteristics of some of the species' local populations. For example, the Chiffchaff, which has an important over-wintering population in Midi-Pyrenees, is found in winter near to stagnant and running waters in the plain, where birds are sedentary. It is therefore strongly affected by rigorous winters, leading to a drop in numbers in riparian forest, observed

between 1985 and 1987 (Lauga & Joachim, 1992). However, local populations, which are mainly migratory outside of riparian forests, are not greatly affected (Affre, 1975*b* ; Joachim *et al*., 1997). Another example implies the European Robin. It was a common bird in this region before the cold January of 1985. Rarer in 1985 spring, it almost disappeared from the study area by spring 1987. It was only present in the largest terrace forest where its frequency was greatly reduced (F = 30% instead of 80% in 1985). This local extinction happened in plains and hills of the piemont Garonne. However, in mountainous areas such as Haute-Garonne Pyrenees and Ariège, no significant frequency variation occurred between 1985 and 1987: F=100% for Cagire beech groves (Haute-Garonne, 1000 m) and beech and birch groves in the Beille plateau (Ariège, 800-1400m) for both years (Joachim *et al*., 1997). Therefore, it can be hypothesized that the European Robin from the Garonne plains was sedentary and did not survive the rigorous winters while the one from Central Pyrenees are migratory and avoided these winters. Génard & Lescourtet (1985) found insular type adaptative characters in European Robin of the Pyrenees. However, this species general range skirts Mediterranean areas and shows no interruption between Pyrenees and the rest of France. Therefore, the authors could not explain this phenomenon. Recent events suggest a distinction between East and Central Pyrenees populations and more northern populations. First, during cold winters, the Garonne population disappears and there is a real separation between areas. Second, during mild winters, a sedentary Garonne population of European Robin develops but because it is sedentary, it stays isolated from migratory Pyrenees' European Robins which have then developed specific insular peninsular characteristics.

Interior vs. edge forest - We have not detected significant differences between interior and edge forest, and globally species occupancy rates seemed to vary in the same way in these two sub-areas. No changes of habitat preferences has been detected, but response to climate

changes might be less important in some cases in the interior forest than in the edge forest indicating that actually, the interior forest might have provided a better protection. The extent of this phenomenon may even be more important that what we could see here. In fact, because of the poor sampling rate in the edge forest, the credible intervals of estimates are large leading to a possible underestimation of the protection role played by the inner forest. It is interesting to note that without the use of prior information, the detection of this effect would have been compromised.

On a larger scale, detected variations in frequency are generally associated to colonization or retreat from areas that do not correspond to preferred habitat (defined by sampling from 1985; Joachim, 1986). Largely wide-spread in woods located in this area, the Chiffchaff has partially retreated from riparian forests while wren populations decreased in terraces. Birds that are mainly forest birds have colonized small groves (e.g. Mistle Thrush, Lesser Spotted Woodpecker), and open-field birds have penetrated into forest edges (e.g. Tree Pipit, Whitethroat). These clues suggest that bird species with continental affinities, even if progressing for various reasons, have difficulties in finding their optimal habitat. The local frequency increase with occupation of unusual environments might reflect healthy neighboring continental populations (e.g. in northern France, the Massif Central, or the Pyrenees) with individuals that would settle in environments marginal to our region. Such phenomena have been demonstrated by Askins & Philbrick (1987), and Askins *et al.* (1987) for forest fauna in Connecticut (USA), by Van Dorp & Opdam (1987) in Dutch groves, and by Helle & Järvinen (1986) in northern Finland.

In open-fields near the study area, ornithologists have pointed out the disappearance of some species (*Cisticola juncidis*, Indo-african faunistic type, *Sylvia undata*, Mediterranean faunistic type), but no disappearance has been noticed in woody areas. *Cettia cetti* has undergone a drastic regression in riparian forests without disappearing, and so has *Erithacus*

rubecula in all forests. Globally, species from forest areas seem to cope better with climatic variations (even if affected) than sedentary open-field species. Some forest species have even profited of new climatic conditions.

Role of the species' biogeographical origins - Continentalization of the regional forest fauna has not been yet reflected by loss or gain of species, but instead by population reduction or expansion by locally common species. However, some species have been more affected and show a general reduction (or even extinction) or global colonization in the studied area. Some extant monitoring programs such as the European Bird Census Council (http://www.ebcc.info/) are interested in the expansion or regression of occupied areas of several species, but no particular attention is drawn to the biogeographical origins of species, which, as shown here, can be crucial.

In a broader context, during the 1970's, a different situation occurred with the settlement or expansion of Mediterranean species (e.g. Dartford Warbler, Subalpine Warbler, Woodchat Shrike), Turkestano-Mediterranean species (e.g. : Cetti's Warbler, Bee-eater) or Indo-African species (Zitting Cisticola) in our study area during a succession of mild winters (Affre 1975*a* and *b*, Yeatman 1974). This expansion often extended to the French Atlantic coast which provided a mild winter climate (Yeatman, 1976). Highly contrasting seasons from 1985, 1986 and 1987 have been correlated with regression and even local extinction of sedentary populations to the benefit of these colonizing species, leaving small populations on the Atlantic coast (e.g. Warblers). Depending on annual climatic variations, the Val de Garonne area is alternatively colonized by Mediterranean thermophilic species using early stages of plant successions, or continental species (especially forest species) uncomfortable in isolated wood patches present in the agricultural plains (Blondel, 1984). These temporary colonizations still permit the exchange of individuals between populations that are

intermittently isolated from each other (e.g. populations from the Pyrenees' continental fauna and mid-Europe reservoir, and Mediterranean populations from the Atlantic coast and Mediterranean reservoir). European partial migratory species with a sedentary edge area that encompass the South-West of France hardly succeed to sustaining populations.

Various biogeographical influences acting on this region, usually thought to be an enriching factor to natural environment, here, are paradoxically related to a noticeable faunistic deficit (Balent *et al.*, 1988). Climatic variations and advanced deforestation only allow commonly wide-spread species and pioneer species. Riparian Garonne forest seems to be important, because regardless of climatic conditions, it contains a fauna of diverse biogeographical regions, such as the Turkestano-Mediterranean Cetti's Warbler , the European Garden Warbler (Affre, 1980) and the palearctic Marsh Tit (Joachim, 1987). Avifauna responses to climatic variations are slower here than in the open-fields of neighboring agricultural plains. While the Garonne riparian forest, with a small area, can not be a local fauna local reservoir and populations therein are not as unchanging as in large forest areas (Glowacinski, 1981). It constitutes a moderating element in Garonne landscape because of its slow response to climatic variations,. It also provides a more or less hospitable link between remote patches of biogeographically different faunas and their source areas.

Taking into account some prior on the detectability of the target species can significantly improve the estimation of its occupancy rates when it has been very few detected (or not detected) and is known to be easily detectable (as already pointed out by Dupuis et al., 2010). As a matter of fact, ignoring such an available prior may have led to different conclusions such as detection of a variation in occupancy rate while not true (e.g. Green Woodpecker or Whitethroat in edge forest), a wrong estimation of the impact of rigorous winters (e.g.

Nightingale or European Robin in edge forest), or worse, it would have failed to detect a diminution of the occupancy (e.g. Nightingale in the inner forest). Moreover, in our dataset, we have undetected species which were known to be present in the study area; we have thus adopted the conditional approach of Dupuis *et al.* (2010) for estimating their occupancy rates. As shown in the comparison between unconditional and conditional approaches, not accounting for the known presence of an undetected species can lead to significant underestimation of its occupancy rate.

Acknowledgements

Authors are grateful to Michael C. Runge, Roger Pradel and Kyle G. Dexter for their helpful comments.

References

Ackerly D.D., 2003. Community assembly, niche conservatism, and adaptative evolution in changing environments. *International Journal of Plant Sciences*. **164(Suppl.)**:165–184.

Affre G., 1975a. Dénombrement et distribution géographique des fauvettes du genre Sylvia dans une région du midi de la France. II. Résultats. *Alauda*. **43**: 229-262.

Affre G., 1975b. Estimation de l'évolution quantitative des populations aviennes dans une région du midi de la France au cours de la dernière décennie (1963-1972). *L'oiseau et la R.F.O*. **45**: 165-187.

Affre G., 1980. Distribution altitudinale des oiseaux dans l'Est des Pyrénées françaises. *L'Oiseau et la R.F.O*. **50**: 1-22.

Ainley D.G., Ballard G., Emslie S.D., Fraser W.R., Wilson P.R. & Woehler E.J., 2003. Adelie penguins and environmental change. *Science*. **300**: 429–430.

Askins R. A. & Philbrick M. J., 1987. Effect of changes in regional forest abundance on the decline and recovery of a forest bird community. *The Wilson bulletin*. **99**: 7-21.

Askins R. A., Philbrick M. J. & Sugeno D. S., 1987. Relationship between the regional abundance of forest and the composition of forest bird communities. *Biological conservation*. **39**: 129-152.

Balda R. P., Gaud W. S. & Brawn J. D., 1983. Predictive models for snag nesting birds. 216-222. *In* J. W. Davis, G. A. Goodwin, and R. A. Ockenfels, Snag Habitat Management : *Proceedings of the Symposium*. U.S. Department of Agriculture, Forest Service, General Technical Report RM-99. Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

Balent G., Génard M. & Lescouret F., 1988. Analyse des patrons de répartition des oiseaux nicheurs en Midi-Pyrénées. *Acta oecologica. Oecologia generalis*. **9**: 247-263.

Blondel J., 1975. L'analyse des peuplements d'oiseaux, éléments d'un diagnostic écologique 1; La méthode des échantillonnages fréquentiels progressifs (E.F.P.). *Revue d'Ecologie*. **29**: 533–589.

Blondel J., 1984. Avifaunes méditerranéennes, histoire des peuplements. *Aves*. **21**: 209-226.

Blotzheim U.N. & Bauer K.M., 1985, 1988, 1993, 1997. *Handbuch der Vogel Mitteleuropas*. Aula-Verlag, Wiesbaden.

Colombo A.E., Etkin D. & Karney B.W., 1999. Climate Variability and the Frequency of Extreme Temperature Events for Nine Sites across Canada Implications for Power Usage. *American Meteorological Society*. **12**: 2490—2502.

Coope G.R., 1994. The response of insect faunas to glacial-interglacial climatic fluctuations. *Philosophical Transactions: Biological Sciences*. **344**: 19–26.

Croxall J.P., Trathan P.N. & Murphy E.J., 2002. Environmental change and Antarctic seabird populations. *Science*. **297**: 1510–1514.

Davis M.B. & Zabinski C., 1992. Changes in geographical range resulting from green-housewarming: effects on biodiversity in forests. 297–308. *In:* Global Warming and Biological Diversity, ed. TEL Peters, R Lovejoy. New Haven, CT: Yale Univ. Press.

Décamps H., Joachim J., & Lauga J., 1987. The importance for birds of the riparian woodlands within the alluvial corridor of the river Garonne, SW France. *Regulated rivers : research and management*. **1**: 301-316.

Derocher A.E., 2005. Population ecology of polar bears at Svarlbad, Norway. *Population Ecology*. **47**: 267–275.

Dupuis J., Bled F. & Joachim J., 2010. Estimating the occupancy rate of spatially rare or hard to detect species: a conditional approach. *Biometrics*. (*in press*)

Dupuis J. & Joachim J., 2006. Bayesian estimation of species richness from quadrat sampling data in the presence of prior information. *Biometrics* **62**, 706-712.

Génard M. & Lescouret F., 1985. Caractères insulaires de l'avifaune forestière des Alpes du sud et des Pyrénées-Orientales. *Acta oecologica. Oecologia generalis.* **6**: 209-221.

Glowacinski Z., 1981. Stability in bird communities during the secondary succession of a forest ecosystem. *Ekologia Polska*. **29**: 73-95.

Grabherr G., Gottfried M. & Pauli H., 1994. Climate effects on mountain plants. *Nature*. **369**: 448.

Grzybowsky J. A., 1983. Patterns of space use in grassland bird communities in winter. *The Wilson Bulletin*. **95**: 591-602.

Hanski I., 1997. Metapopulation dynamics: from concepts and observations to predictive models. 69-91. *In*: Metapopulation biology. Eds.: I. Hanski & M. E. Gilpin. San Diego: Academic Press.

Hejl S. J. & Beedy E. C., 1986. Weather induced variation in the abundance of birds. 241-244. *In*: Verner J., Morisson M. L. and Ralph C. J. Wildlife 2000, Modeling Habitat

Relationships of Terrestrial Vertebrates. The University of Wisconsin Press, Madison, Wisconsin.

Helle P. & Järvinen O., 1986. Population trends of north finnish land birds in relation to their habitat selection and changes in forest structure. *Oïkos*. **46**: 107-115.

Hill J.K., Thomas C.D., Fox R., Telfer M.G. & Willis S.G., 2002. Responses of butterflies to twentieth century climate warming: implications for future ranges. *Proceedings of the Royal Society B.* **269**: 2163–2171.

Hoegh-Guldberg O., 1999. Climate change, coral bleaching and the future of the world's coral reefs. *Marine & Freshwater Research*. **50**: 839–866.

Huntley B., 1991. How plants respond to climate change: migration rates, individualism and the consequences for the plant communities. *Annals of Botany*. **67**: 15–22.

Joachim J. & Lauga J., 1992. Influence des hasards climatiques sur les communautés de passereaux nicheurs dans les boisements du couloir alluvial garonnais. *Le Pistrac*. **14**: 1-17. Muséum Toulouse.

Joachim J., 1986. Influence du morcellement forestier sur les peuplements d'oiseaux nicheurs dans le couloir alluvial garonnais. *Ph.D. thesis in Ecology*. Université Toulouse III.

Joachim J., 1987. La Mésange nonnette *Parus palustris* dans les ripisilves garonnaises. *Alauda*. **55**: 112-115.

Joachim J., Bousquet J.F. & Fauré C., 1997. Atlas des oiseaux nicheurs de Midi-Pyrénées. AROMP *Ed.* Museum d'Histoire Naturelle, Toulouse. 262p.

Kendall W. L., 2001. Using models to facilitate complex decisions. 147–170 *In*: Modeling in natural resource management. *Eds:* T. M. Shenk & A. B. Franklin. Island Press, Washington, D.C.

Krebs C., 1978. Ecology: The Experimental Analysis of Distribution and Abundance. Harper and Row Publishers, Inc., New York, New York, USA.

Lande R., 1988. Demographic models of the northern spotted owl (Strix occidentalis caurina). *Oecologia*. **75**: 601-607.

Lauga J. & Joachim J., 1992. Modelling the effects of forest fragmentation on certain species of forest-breeding birds. *Landscape Ecology*. **6(3)**: 183-193.

Levins R., 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*. **15**: 237-240.

Levins R., 1970. Extinction. 75-107 *In*: Some mathematical problems in biology. Ed: M. Gerstenhaber. Providence, RI: American Mathematical Society.

MacKenzie D.I., Nichols J.D., Lachman G.B., Droege S., Royle A.J. & Langtimm A.C., 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*. **83**: 2248-2255.

MacKenzie D.I., Nichols J.D., Royle J.A., Pollock K.H., Bailey L.L. & Hines J.E., 2006. *Occupancy Estimation and Modeling*. Academic Press. San Diego, California, USA. 324p.

Malcom J.R., Liu C., Neilson R., Hansen L. & Hannah L., 2006. Global Warming and Extinctions of Endemic Species from Biodiversity Hotspots. *Conservation Biology*. **20(2)**: 538–548.

Mikkola K., 1997. Population trends of Finnish Lepidoptera during 1961–1996. *Entomologica Fennica.* **3**: 121–143.

Nichols J.D., 2004. Evolution of quantitative methods for the study and management of avian populations: on the importance of individual contributions. *Animal Biodiversity and Conservation*. **27(1)**: 3–19.

Parmesan C., 2006. Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*. **37**: 637–69.

Parmesan C., Ryrholm N., Stefanescu C., Hill J.K. & Thomas C.D., 1999. Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature*. **399**: 579–583.

Pauli H., Gottfried M. & Grabherr G., 1996. Effects of climate change on mountain ecosystems: upward shifting of mountain plants. *World Resource Review*. **8**: 382–390.

Pounds J.A., Fogden M.P.L. & Campbell J.H., 1999. Biological response to climate change on a tropical mountain. *Nature*. **398**: 611–615.

Pounds J.A., Fogden M.P.L. & Masters K.L., 2005. Case study: responses of natural communities to climate change in a highland forest. 70-74. *In:* Climate change and biodiversity. *Eds*: T.E. Lovejoy & L. Hannah. Yale University Press, London.

Spitz F., 1974. Facteurs de répartition de l'avifaune en forêt de montagne. *In*: Écologie forestière: la forêt: son climat, son sol, ses arbres, sa faune: ouvrage collectif. *Ed*: Pearson P. Gauthier-Villars, Paris. 382 p.

Stirling I., Lunn N.J. & Iacozza J., 1999. Long-term trends in the population ecology of polar bears in western Hudson Bay in relation to climatic change. *Arctic*. **52**: 294–306.

Thomas C.D. & Lennon J.J., 1999. Birds extend their ranges northwards. *Nature*. **399**: 213.

van Dorp D. & Opdam P.F.M., 1987. Effects of patch size, isolation and regional abundance on forest bird communities. *Landscape Ecology*. **1**: 59-73.

Wiens J.A., 1981. Single-sample surveys of communities : are the revealed patterns real? *American naturalist*. **117**: 90-98.

Wilkinson C.R., ed., 2000. Global Coral Reef Monitoring Network: Status of Coral Reefs of the World in 2000. Townsville, Qld: Aust. Inst. Mar. Sci

Williams B.K., Nichols J.D. & Conroy M.J., 2002. Analysis and Management of Animal Populations. *San Diego, California, Academic Press*.

Yeatman L., 1974. Les rapports entre des caractères climatiques et botaniques et la distribution des oiseaux méditerranéens en France. *L'oiseau et la R.F.O.* **44**: 324-339.

Yeatman L., 1976. Atlas des oiseaux nicheurs de France. Société Ornithologique de France, Paris.

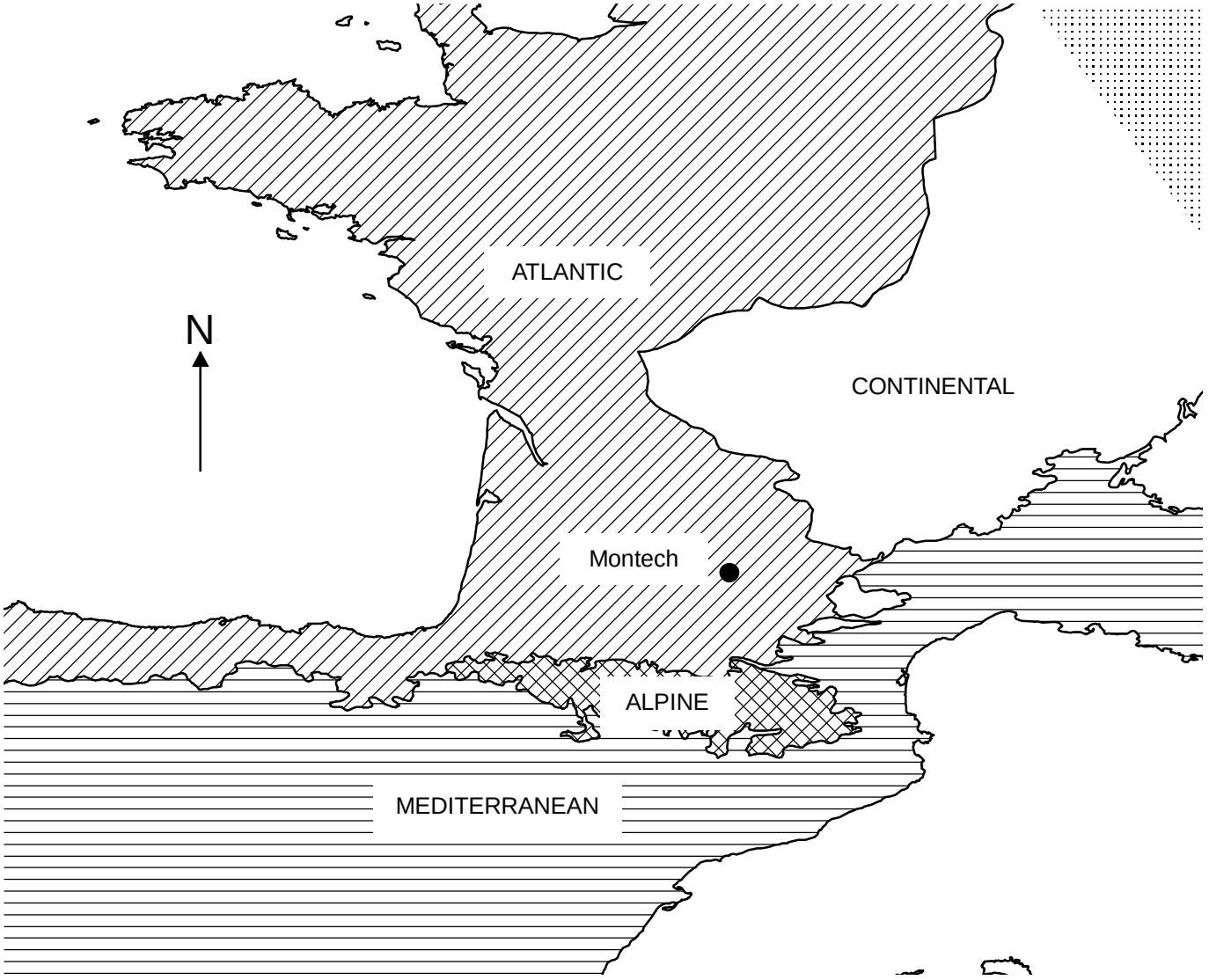
Tables and Figures

Figure 1. European map of biogeographical region and location of Montech Forest. (using data provided by the European Environmental Agency: www.eea.eu.int*). (Black point: Montech Forest, Cross hatch: Alpine, Diagonal simple hatch: Atlantic, Ordered stipple: Continental, Horizontal simple hatch: Mediterranean).*

Figure 2. Mean monthly temperatures in 1985, 1986 and 1987 *(Spotted line: 1985, Hachted line: 1986, solid line: 1987)* compared to 1970-1984 period (maximum and minimum values from this 14 year period bounded in gray).. (from Joachim & Lauga, 1992)

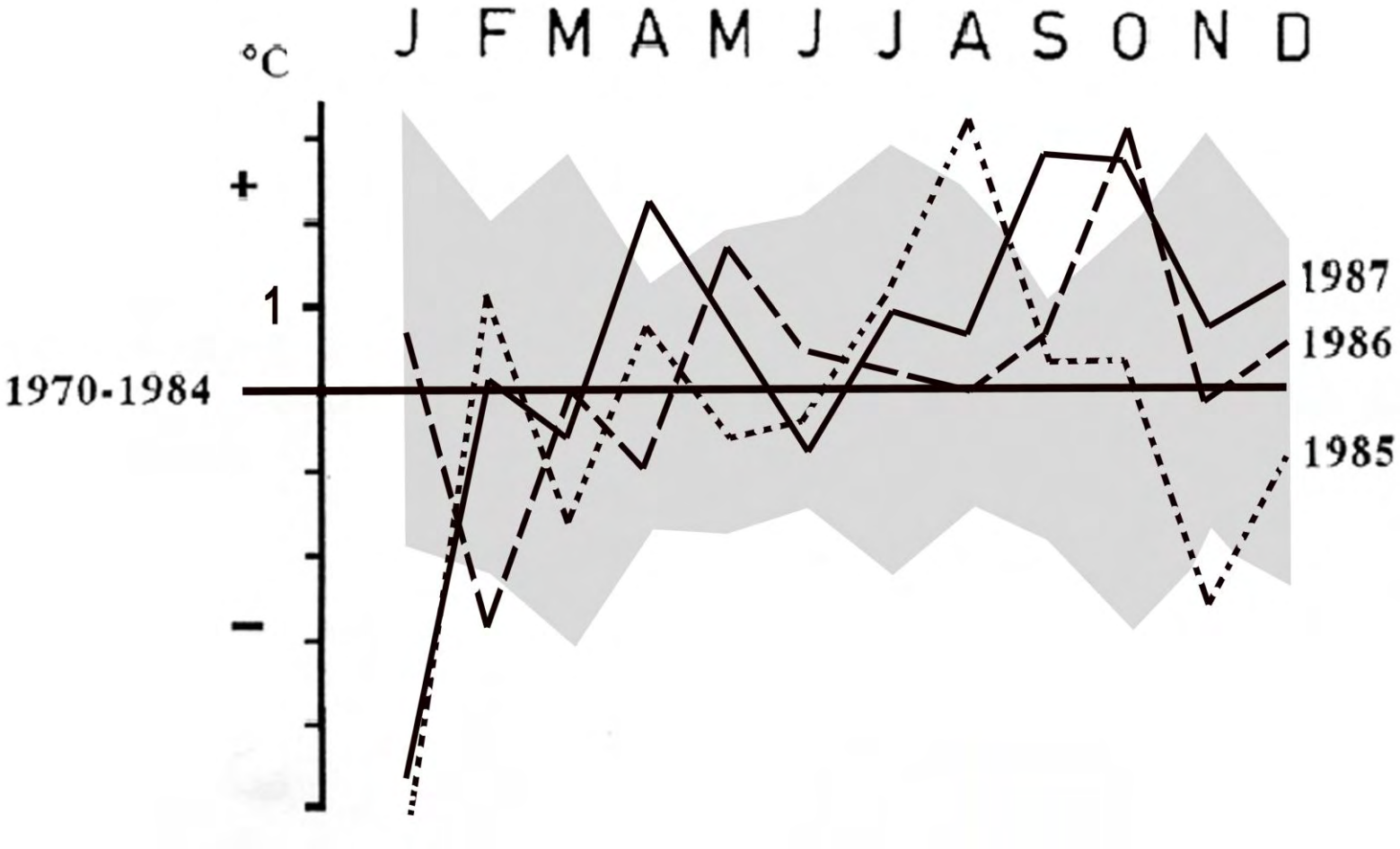
Figure 1.

1



2

Figure 2.



1 Table 1.

2

Table 1. Data ($V_s(a);W_s(a)$) for 1985 and 1987 in Montech forest.

3

Species		Inner forest		Edge forest	
4		1985	1987	1985	1987
Song Thrush	*Turdus philomelos*	(20 ; 53)	(10 ; 20)	(16 ; 43)	(1 ; 2)
Common Blackbird	*Turdus merula*	(17 ; 32)	(12 ; 25)	(14 ; 35)	(4 ; 6)
Whitethroat	*Sylvia communis*	-	-	(2 ; 2)	(1 ; 1)
Chiffchaff	*Phylloscopus collybita*	(15 ; 30)	(5 ; 13)	(10 ; 24)	(4 ; 8)
Eurasian Nuthatch	*Sitta europea*	(7 ; 17)	(9 ; 26)	(4 ; 8)	(2 ; 2)
Green Woodpecker	*Picus viridis*	(1 ; 1)	(3 ; 3)	(0 ; 0)	(2:2)
Short-toed Treecreeper	*Certhia brachydactyla*	(15 ; 22)	(10 ; 17)	(11 ; 28)	(2 ; 2)
European Robin	*Erithacus rubecula*	(9 ; 17)	(5 ; 8)	(7 ; 12)	(2 ; 6)
Nightingale	*Luscinia megarhynchos*	(4 ; 9)	(1 ; 1)	(5 ; 7)	(0 ; 0)
Mistle Thrush	*Turdus viscivorus*	(6 ; 13)	(8 ; 22)	(4 ; 5)	(3 ; 4)

1 Table 2.

2

Table 2. Migratory status and biogeographical origins of the selected species. *Migratory status: Sedentary (S), Migrant (M) or Partial migrant (P).*

3

Species	Migratory status	Global faunistic class	European biogeographical origin
Song Thrush	P	European	Atlantico-continental
Common Blackbird	P	Palearctic	European
Whitethroat	M	Turkestano-european	Atlantico-mediterranean
Chiffchaff	M	Palearctic	Atlantico-continental
Eurasian Nuthatch	S	Palearctic	European
Green Woodpecker	S	European	European
Short-toed Treecreeper	S	European	European
European Robin	P	European	Atlantico-continental
Nightingale	M	European	Atlantico-mediterranean
Mistle Thrush	P	European	Continental

1 Table 3.

2

Table 3. Prior mean and CI $_{95\%}$ of detection probability μ_{sa} during 1985 and 1987 in Montech forest.

Species	Inner forest		Edge forest	
	1985	1987	1985	1987
Song Thrush	0.8 [0.65 – 1]	0.6 [0.3 – 0.8]	0.8 [0.65 – 1]	0.4 [0 – 0.65]
Common Blackbird	0.8 [0.65 – 1]	0.6 [0.3 – 0.8]	0.8 [0.65 – 1]	0.8 [0.65 – 1]
Whitethroat	-	-	0.8 [0.65 – 1]	0.8 [0.65 – 1]
Chiffchaff	0.8 [0.65 – 1]	0.8 [0.65 – 1]	0.8 [0.65 – 1]	0.8 [0.65 – 1]
Eurasian Nuthatch	0.8 [0.65 – 1]	0.8 [0.65 – 1]	0.6 [0.3 – 0.8]	0.6 [0.3 – 0.8]
Green Woodpecker	0.6 [0.3 – 0.8]	0.6 [0.3 – 0.8]	0.8 [0.65 – 1]	0.8 [0.65 – 1]
Short-toed Treecreeper	0.8 [0.65 – 1]	0.8 [0.65 – 1]	0.6 [0.3 – 0.8]	0.6 [0.3 – 0.8]
European Robin	0.8 [0.65 – 1]	0.6 [0.3 – 0.8]	0.8 [0.65 – 1]	0.2 [0 – 0.4]
Nightingale	0.6 [0.3 – 0.8]	0.6 [0.3 – 0.8]	0.8 [0.65 – 1]	0.8 [0.65 – 1]
Mistle Thrush	0.6 [0.3 – 0.8]	0.6 [0.3 – 0.8]	0.4 [0 – 0.65]	0.4 [0 – 0.65]

Table 4.

Table 4. Posterior mean and CI $_{95\%}$ of γ_{sa} for inner and edge regions of Montech forest in 1985 and 1987. Estimates were computed using informative prior on q_{sa}.

	Inner forest				Edge forest			
Species	1985		1987		1985		1987	
Song Thrush	0.95	[0.91 ; 1]	0.85	[0.68 ; 1]	0.93	[0.89 ; 1]	0.47	[0.06 ; 0.94]
Common Blackbird	0.89	[0.77 ; 1]	0.93	[0.77 ; 1]	0.84	[0.78 ; 0.94]	0.76	[0.44 ; 1]
Whitethroat		-		-	0.16	[0.11 ; 0.28]	0.08	[0.06 ; 0.17]
Chiffchaff	0.79	[0.68 ; 0.96]	0.42	[0.23 ; 0.59]	0.63	[0.56 ; 0.78]	0.74	[0.44 ; 1]
Eurasian Nuthatch	0.37	[0.32 ; 0.45]	0.68	[0.50 ; 0.86]	0.36	[0.22 ; 0.61]	0.59	[0.22 ; 1]
Green Woodpecker	0.11	[0.05 ; 0.32]	0.46	[0.14 ; 0.86]	0.07	[0.05 ; 0.14]	0.49	[0.11 ; 0.83]
Short-toed Treecreeper	0.87	[0.73 ; 1]	0.84	[0.64 ; 1]	0.72	[0.61 ; 0.89]	0.59	[0.11 ; 0.94]
European Robin	0.51	[0.41 ; 0.68]	0.60	[0.32 ; 0.96]	0.50	[0.39 ; 0.72]	0.57	[0.11 ; 0.94]
Nightingale	0.28	[0.18 ; 0.45]	0.19	[0.05 ; 0.50]	0.39	[0.28 ; 0.61]	0.15	[0.06 ; 0.39]
Mistle Thrush	0.39	[0.27 ; 0.59]	0.67	[0.45 ; 0.91]	0.59	[0.28 ; 1]	0.75	[0.39 ; 1]

Table 5.

Table 5. Posterior mean of γ_{sa} and CI $_{95\%}$ for inner and edge regions of Montech forest in 1985 and 1987. Estimates were computed using a non informative prior on q_{sa}. *Only results differing from the informative approach are presented here.*

	Inner forest				Edge forest			
Species	1985		1987		1985		1987	
Song Thrush	-		-		-		0.29	[0.06 ; 0.72]
Common Blackbird	-		-		-		-	
Whitethroat	-		-		0.41	[0.11 ; 0.94]	0.42	[0.06 ; 0.94]
Chiffchaff	-		-		-		-	
Eurasian Nuthatch	-		-		-		-	
Green Woodpecker	0.24	[0.05 ; 0.82]	0.56	[0.23 ; 1]	0.25	[0.05 ; 0.82]	0.59	[0.22 ; 1]
Short-toed Treecreeper	-		-		-		-	
European Robin	-		0.52	[0.23 ; 0.86]	-		0.37	[0.11 ; 0.61]
Nightingale	-		0.42	[0.06 ; 0.94]	-		0.24	[0.05 ; 0.78]
Mistle Thrush	-		0.58	[0.36 ; 0.73]	-		0.67	[0.33 ; 1]

Table 6a. Comparison between γ_{sa} estimates under unconditional (MacKenzie *et al.*, 2006) and conditional (Dupuis *et al.*, 2010) approaches when a species is not detected, for a non informative and an informative prior on q_{sa}: case of the Green Woodpecker in edge region of Montech forest in 1985. *J=T=18.*

	Unconditional approach		Conditional approach	
Non informative prior	0.13	[0 ; 0.68]	0.25	[0.05 ; 0.82]
Informative prior	0.02	[0 ; 0.11]	0.07	[0.05 ; 0.14]

Table 6b. Comparison between γ_{sa} estimates under unconditional (MacKenzie *et al.*, 2006) and conditional (Dupuis *et al.*, 2010) approaches when a species is not detected, for a non informative and an informative prior on q_{sa}: case of the Nightingale in edge region of Montech forest in 1987. *J=18 , T=6.*

	Unconditional approach		Conditional approach	
Non informative prior	0.20	[0 ; 0.78]	0.24	[0.05 ; 0.78]
Informative prior	0.12	[0 ; 0.39]	0.15	[0.05 ; 0.39]

CHAPTER VI

CONCLUSION

In this thesis, I have addressed -especially in the framework of hierarchical modeling- several of the main limitations occupancy models are reproached with (e.g. MacKenzie *et al*., 2005 ; Royle & Dorazio, 2008), such as the lack of spatiotemporal models and explicit/mechanistic formulation of occupancy, and detectability problems. I have also identified issues that might emerge under some circumstances with conditioning assumptions about occupancy.

In this last chapter, I will discuss the choice of using the hierarchical modeling approach to assess site occupancy, and why I decided to focus on site occupancy rather than on abundance. Finally, I will highlight several developments or improvements that can be considered in order to enhance the already large framework of hierarchical modeling of site occupancy.

Hierarchical modeling

In this thesis, site occupancy is mainly treated from a hierarchical modeling point of view since accounting for uncertainty in observation was one of my main objectives. Indeed, explicit consideration of detectability is essential when occurrence-based summaries of population status are the focus of inference (Royle & Dorazio, 2006). Here, hierarchical modeling refers to models that make a clear distinction between the observation component and the process component. Moreover, even if recently it has become synonymous to Bayesian analysis, and even if this thesis is placed in a Bayesian context, hierarchical models do not have to be estimated using a Bayesian approach.

I have developed in chapter II a hierarchical model for an invasion process and specified more clearly the underlying assumptions that can be made when modeling site occupancy under the presence of uncertainty in chapters IV and V. The hierarchical formulation of site occupancy models yield to a generic, flexible, and practical framework for modeling individual ('site') effects, or other latent structure in parameters (e.g. random year effects) in dynamic occupancy models. Models with explicit spatial dynamics are necessarily formulated as individual effects models, where the individual effect is a function of occupancy states at nearby spatial units, leading to models that cannot have a "non-hierarchical" formulation in some cases (Royle & Dorazio, 2008).

Moreover, hierarchical models provide a general and flexible framework for addressing inference problems about animal occurrence from spatially referenced survey data that are subject to imperfect detection, and they also permit a unified treatment of modeling and inference for a vast array of animal sampling methods (Royle & Dorazio, 2006). The hierarchical modeling framework can be applied in a large and diverse set of problems in ecology –estimating occupancy for a single species, estimating population size in a single closed population, estimating community structure, etc… (Royle & Dorazio, 2008). Although a number of non-Bayesian methods of inference of animal occurrence in the context of imperfect detection can be described (see, e.g. Kéry *et al.*, 2005 ; Dorazio *et al.*, 2005), the lack of generality of such methods contrasts with the flexibility and rigor of the hierarchical Bayesian formulation (Royle & Dorazio, 2006). The hierarchical framework provides a natural framework for developing models for spatio-temporal dynamics (Wikle, 2003) and multiple species systems (Dorazio & Royle, 2005).

For all the above reasons, hierarchical modeling appeared to be a logical and reasonable way of approaching site occupancy modeling in this thesis, and while it may still have some flaws (such as evocated in chapter V), it already provides a reliable framework to model site occupancy, and abundance.

Occupancy versus abundance: different costs for the same inferences?

Knowing how individuals or species occupy their habitat is essential to understand how species use their environment (Krebs, 1978), and take suitable management dispositions (Kendall, 2001; Williams *et al.*, 2002 ; Nichols, 2004 ; MacKenzie *et al.*, 2005). Investigations of animal populations often require estimates of animal abundance or occurrence. These estimates may be used to monitor spatial or temporal changes in the population or to determine whether relationships exist between animals and one or more environmental characteristics (i.e., to identify habitat) (Royle & Dorazio, 2006). As previously said, abundance can be estimated from occupancy data. For example, if individuals are distributed randomly over space, their distribution will follow a Poisson distribution which can be computed (MacKenzie *et al.*, 2005). This approach has a long history in plant and animal ecology (e.g. Gleason, 1920 ; Fisher, 1922). Of course, depending on the

relationship between abundance and occupancy, other distributions can be preferred such as the negative binomial (e.g. Fisher, 1922 ; Dice, 1948 ; Gerrard & Chaing, 1970). A more recent example of dealing with abundance estimation using occupancy data can be found in He & Gaston (2000, 2003) through the combination of phenomenological models of occupancy-abundance and mean-variance relationships. Occupancy status is often is significantly less costly to obtain and simpler to collect, compared to abundance data,. However, this comes at a price, information content has to be compromised for this efficiency and economy. When one decides to design a sampling protocol, this compromise has to be kept in mind, and future development of site occupancy model surely includes the improvement of abundance estimation with detection/non-detection data.

In the specific case of a spatially closed environment, site occupancy rate is also one of the variables that are useful to determine how species use their environment. While density and abundance carry information on the population state, occupancy rate describes how species use the environment they live in. It is defined as the proportion of patches occupied and is viewed as a state variable in various metapopulations models (e.g., Levins, 1969, 1970 ; Lande, 1987, 1988 ; Hanski, 1992, 1994, 1997). By definition, site occupancy rate has to be considered in a 'closed environment', i.e. a finite population of locations called sites or quadrats. While this has been well studied for large sites population, improvements such as presented in chapters IV and V can be made in the case of smaller sites population (MacKenzie *et al.*, 2002). Moreover, the distinction between probability of occupancy and the proportion of sites becomes more important with the decrease in sites population size (MacKenzie *et al.*, 2005). Site occupancy rate is especially useful in wildlife management and endangered species conservation programs where controlled areas often are limited but can be fully sampled (e.g. a network of ponds or forests in a park) (MacKenzie *et al.*, 2005). It permits to know where actions have to be taken and how these actions will impact species repartition. Moreover the link between species and their habitat can be used to characterize the specificities of this environment. Using a list of species of interest, one can use these species site occupancy rates to determine the species composition characteristic of particular ecoregions or ecosystems. An application would be the use of bioindicators to assess the evolution of environmental and ecological quality of an area (Kolkwitz & Marsson, 1902 *in* van Straalen, 1998 ; Burgeot *et al.*, 1996 ; Bongers & Ferris, 1999).

How to improve estimates when modeling site occupancy: incorporation of scientific insights and sampling design.

Incorporating scientific insights – Ecological processes exhibit complicated behavior over an extensive range of spatial and temporal scales of variability. To understand and eventually predict such complicated processes, we must make use of available scientific insight, data, and theory, in a modeling framework that honestly accounts for uncertainties in each of them (Wikle, 2003). The question is 'how to incorporate scientific insight into the hierarchical framework?'. For example, it is possible to incorporate inexact scientific theory into the hierarchical framework, such as suggested by deterministic models, but that proves to be useful (e.g. Wilkle *et al.*, 2001).

Ecological prediction is difficult. It is even more so in the presence of uncertainty. Observations have errors and we seldom know the true underlying dynamical model for a given phenomenon. For example, in the context of invasive species, there may be demographic and environmental stochasticities, Allee dynamics, coalescing colonies, jump diffusions and heterogeneous rates of spread and growth. Theoretical models, although able to include some of these effects in simplistic scenarios, generally do not have the flexibility to accommodate the a priori uncertainty related to the dynamical assumptions. On the other hand purely stochastic models for such processes are often overparameterized and face significant problems when it comes to estimation (Wikle, 2003). The Bayesian hierarchical framework can be used to build complicated spatiotemporal prediction models and quantify the associated uncertainties. However, one of the issues that arise when using a Bayesian hierarchical model is how to decide on the appropriate parameterizations. Often, there is scientific insight for the most critical processes. One seeks to allow the models that are based on this insight to have the flexibility to fit to the data. Moreover, multiple parameterizations might work equally well (Wikle, 2003).

Hooten & Wikle (2008) have implemented an ecologically meaningful Partial Differential Equation within a hierarchical Bayesian framework as a latent dynamical system to help manage such uncertainty and account for complicated dependence structures in parameters. They have used a reaction-diffusion model for the latent process to represent scientific opinion. Diffusion models have long been considered for describing the spread of invasive

organisms, and they have relevance in many invasive bird species (e.g., Okubo, 1986 ; Veit & Lewis, 1996). But it has to be specified that it is not the only way of incorporating scientific knowledge into a model. For example, matrix models (e.g. Caswell, 2001) may prove to be a promising tool in future invasive species research efforts. Such settings allow for the inclusion of well-known non-linear growth and dispersal equations acting upon a more intuitive latent process. Implementation of these models, however, presents a variety of challenges (Hooten & Wikle, 2008).

In chapter IV, presence and detectability probabilities (i.e. φ and q) are used to determine occupancy rates. In order to improve estimators performances, several elements can be modified in the sampling design, and in statistical analyses.

Impact of A priori information – Another way of incorporating scientific knowledge is to use *a priori* on detectability. Prior information on detectability is a commonly available resource given by experts or prior studies for example. Such information might indirectly affect the quality of $\hat{\varphi}$ even if not precise. As shown in the study presented in the complementary work of chapter IV, and in the application paper in chapter V, the incorporation of *a priori*, even if it is not on the parameter of interest (i.e. occupancy rate) but on the nuisance parameter (i.e. detection probability), can greatly improve the quality of estimators. This is true even if the a priori is biased or not very precise. Therefore, when available, this type of information should be taken into account, and this is especially easy to do in the Bayesian framework.

Sampling design – Challenging aspects of occurrence estimation include the presence of significant observational error (Sauer *et al.*, 1994), irregular spatial and temporal sampling, the diffusive nature of invasive processes on landscape scales (Wikle, 2003). Several protocols have been conceived in order to estimate occurrence or abundance of animals and their probabilities of detection simultaneously. These protocols include distance sampling, sampling with multiple observers, capture-recapture and removal sampling. (Williams *et al.*, 2002). As shown in chapter IV, one way of improving estimation of detectability, and therefore potentially presence estimator, is to visit each quadrat repeatedly, and this improvement appears to be pretty fast. I have also shown that sampling a larger portion of the area of study increases estimators' quality; it is possible to obtain good estimations of presence and detectability parameters without having to sample the whole study area.

MacKenzie *et al.* (2005) have pointed out that a number of practical options exist to conduct repeated surveys. In chapter IV, I considered only the case where one modifies the number of surveys by doing multiple surveys during one visit, or by doing one survey over several visits. Of course, one must not loose sight of questions and prerequisites involved (such as autocorrelation of detection probability) when designing the protocol. Protocol design cannot ignore the scale considered because of the impact of the latter on efforts and costs devoted to the study. Moreover I have pointed out that little modifications can have a great impact on estimators' performances. Results corroborate hypotheses on allocation effort in the number of sites and the number of surveys presented by MacKenzie *et al.* (2005) who have shown that the most appropriate study can be different depending on variation of the true value of φ and q. A poorly designed study may yield no useful information and may only succeed in wasting resources (MacKenzie *et al.*, 2005). With few modifications in a classic study protocol, estimators' performances can be greatly improved, and these modifications do not have to be highly resource consuming.

Future directions

Multiple occupancy states – In the framework of occupancy modeling, occupancy is often treated as a dichotomous state (occupied vs unoccupied). However, in practice, one can be interested in treating site occupancy as a multistate variable. MacKenzie *et al.* (2005) referred at the frequent case of considering simultaneously breeding and occupancy states. Here, the underlying idea is that the contribution of occupied sites to metapopulation dynamics might differ depending on the fact that reproduction occurs or not on these particular sites (e.g. Pulliam, 1988). The way I am treating occupancy status in chapter II and III can be viewed as a surrogated way of dealing with multiple occupancy states using a compound of two variables, one accounting for occupancy, and one for site history. While this is an interesting and flexible way of treating multiple occupancy status, this should be considered as a premise of ways of dealing with such a topic.

Using information from the capture of marked animals – In chapter III I mentioned the problem that emerges from the different scenarios that might lead to the abandon of a site. For example, a site can become unoccupied because the individual (or population) that occupied it moved away, or because of the death of the previous owner (or local population). On the other

hand, site persistence might be the result of site fidelity or immediate recolonization by a different individual. MacKenzie & Nichols (2004) suggested that one source of potentially useful auxiliary information is obtained from having marked individuals in the study population. I have decided to use information from CMR data beside of the site occupancy modeling because the hierarchical model presented in chapter III is already involving a lot of parameters, however the marked individuals effectively provide information about the movement of individuals among sites, hence making possible to identify the exact processes underlying site occupancy status. Therefore, when possible and reasonable, integrating CMR data directly in the hierarchical site occupancy modeling should be considered.

Favoring a multi-disciplinary approach – Space-time modeling has involvements in countless application areas. In this thesis, I have presented numerous ecological topics in the framework of hierarchical space-time modeling of site occupancy, but such work can be extended into several areas such as social sciences, geostatistics or hydrology. There are many important areas where spatio-temporal data are used to detect recognizable and meaningful patterns as well as for predictions (Sahu & Mardia, 2005). Examples include ecology, geology, many areas of medicine such as brain imaging, wildlife population monitoring and tracking, air-pollution monitoring or disease mapping.

For example, in environmental pollution monitoring, space-time modeling has some history (e.g. Guttorp *et al.*, 1994 ; Haas, 1995 ; Carroll *et al.*, 1997). In recent years, hierarchical Bayesian approaches for spatial prediction of air pollution have been developed (Brown *et al.*, 1994). However, a frequent criticism is directed to spatial prediction using air-pollution data from large-scale monitoring network (Sahu *et al.*, 2006). Here, networks were designed to capture peak pollution levels within urban, highly populated areas. This leads to a potential over-prediction within sparsely monitored rural areas with misleading prediction errors (Sahu & Mardia, 2005). A hierarchical space-time model has been proposed by Sahu *et al.*, 2006. It introduced two spatio-temporal processes (one for rural or background effects, and one adding extravariability for urban/suburban locations). However, the hierarchical model including heterogeneity in the environment we have proposed in the chapter V could be used with few adjustments and a clear formulation of a priori hypotheses.

Other examples of hierarchical space-time modeling include climate and environmental topics such as the spatio-temporal modeling of rainfall and precipitation. This area is currently receiving a great deal of attention. Recent articles on rainfall modeling include Brown *et al.* (2001), Sansó & Guenni (1999, 2000), Allcroft & Glasbey (2003), Sahu *et al.* (2005). In

archaeology, particular modifications to standard models should be done because of the presence of uncertainty on both temporal and spatial scales due to dating imprecision and modification of landscapes (Sahu & Mardia, 2005). Blackwell & Buck (2003) illustrated the use of a fully Bayesian model in which temporal uncertainties are formally accounted for but the spatial information is not explicitly modeled at all. The way of describing the spatial structure in chapter III and IV could be used jointly to the model proposed by Blackwell & Buck (2003) including a weight parameter matrix between sites in order to specify or estimate uncertainty on the relationship between all sites (i.e. the spatial structure itself).

Most ecologists and monitoring program managers were slow to acknowledge the issue that arises from uncertainty in detection (Royle & Kéry, 2007). As a consequence, instead of true species richness, raw (uncorrected) numbers of species or occurrence are reported by most monitoring programs (Weber *et al.*, 2004) and also in most biogeography and community ecology research, with notable exceptions (e.g., Doherty *et al.*, 2003 ; Karanth *et al.*, 2006), of course. There is much to do with the exploration of hierarchical spatiotemporal models in ecology and other disciplines (Wikle, 2003). These efforts, by their very nature, will require strong collaborations between applied mathematicians, statisticians, and subject-matter scientists and the future of site occupancy hierarchical modeling can only be productive if interactions between all these actors are enhanced.

References

Allcroft D. J. & Glasbey C.A., 2003. A latent GaussianMarkov random-field model for spatiotemporal rainfall disaggregation. *Applied Statistics.* **52**: 487–498.

Blackwell P.G. & Buck C.E., 2003. The Late Glacial human reoccupation of north western Europe: new approaches to space-timemodelling. *Antiquity*. **296**: 232–240.

Bongers T. & Ferris H., 1999. Nematode assemblage structure as a bioindicator in environmental monitoring. *Trends in Ecology and Evolution*. **14**: 224-228.

Brown P.E., Diggle P.J., Lord M.E. & Young P.C., 2001. Space-time calibration of radar rainfall data. *Applied Statistics*. **50**: 221–241.

Brown P.J., Le N.D. & Zidek J.V., 1994. Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics*. **22**: 489–510.

Burgeot T., Bocquéné G., Porte C., Dimeet J., Santella R.M., Garcia de la Parra L.M., Pfhol-Leszkowicz A., Raoux C. & Galgani F., 1996. Bioindicators of pollutant exposure in the northwestern Mediterranean Sea. *Marine Ecology Progress Series*. **131**: 125-141.

Carroll R.J., Chen R., George E.I., Li T.H., Newton H.J., Schmiediche H. & Wang N., 1997. Ozone exposure and population density in Harris County, Texas. *Journal of the American Statistical Association*. **92**: 392–404.

Caswell H., 2001. Matrix population models: construction, analysis, and interpretation. *Sinaver*. Massachusetts, USA. 328 p.

Dice L.R., 1948. Relationship between frequency index and population density. *Ecology*. **29(3)**: 389—391.

Doherty P.F. Jr., Boulinier T. & Nichols J.D., 2003. Extinction rates at the center and edge of species' ranges. *Annales Zoologici Fennici*. **40**: 145—153.

Dorazio R.M. & Royle J.A., 2005. Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*. **100**: 389–398.

Dorazio R.M., Jelks H.L. & Jordan F., 2005. Improving removed-based estimates of abundance by sampling a population of spatially distinct subpopulations. *Biometrics*. **61**: 1093—1101.

Fisher R.A., 1922. On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*. **222**: 309-368

Gerrard D.J. & Chaing H.C., 1970. Density estimation of corn rootworm egg populations based upon frequency of occurrence. *Ecology*. **51(2)**: 237—245.

Gleason H.A., 1920. Some applications of the quadrat method. *Bulletin of the Torrey Botanical Club*. **47(1)**: 21—33.

Guttorp P., Meiring W. & Sampson P. D., 1994. A Space-time Analysis of Ground-level Ozone Data. *Environmetrics*. **5**: 241–254.

Haas T.C., 1995. Local Prediction of a Spatio-Temporal Process With an Application to Wet Sulfate Deposition. *Journal of the American Statistical Association*. **90**: 1189–1199.

Hanski I., 1992. Inferences from ecological incidence functions. *The American Naturalist*. **139**: 657–662.

Hanski I., 1994. Patch-occupancy dynamics in fragmented landscapes. *Trends in Ecology and Evolution*. **9**: 131—135.

Hanski I., 1997. Metapopulation dynamics: from concepts and observations to predictive models. Pages 69–91 *in* I. A. Hanski and M. E. Gilpin, eds. *Metapopulation biology: ecology, genetics, and evolution*. Academic Press, New York, New York, USA.

He F. & Gaston K.J., 2000. Estimating species abundance from occurrence. *The American Naturalist*. **156**: 553–559.

He F. & Gaston K.J., 2003. Occupancy, Spatial Variance, and the Abundance of Species. *The American Naturalist*. **162(3)**: 366—375.

Hooten M.B. & Wikle C.K., 2008. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environmental and Ecological Statistics*. **15**: 59 – 70.

Karanth K.K., Nichols J.D., Sauer J.R. & Hines J.E., 2006. Comparative dynamics of avian communities across edges and interiors of North American ecoregions. *Journal of Biogeography*. **33**: 674–682.

Kendall W.L., 2001. Using models to facilitate complex decisions. In: *Modeling in Natural Resource Management* (T.M. Shenk & A.B. Franklin, *eds.*). 147 – 170. Island Press, Washington, D.C., USA.

Kéry M., Royle J.A. & Schmid H., 2005. Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications*. **15**:1450–1461.

Kolkwitz R. & Marsson M., 1902. Grundsätze für die biologische Beurtheilung des Wassers nach seiner Flora und Fauna. Mitteilungen aus der Königlichen Prüfungsanstalt für Wasserversorgung und Abwasserbeseitigung zu Berlin. **1**: 33–72.

Krebs J.R., 1978. Optimal foraging: decision rules for predators. *In*: Krebs JR, Davies NB, *eds*. *Behavioural ecology: an evolutionary approach*. Oxford: Blackwell. p 23–63.

Lande R., 1987. Extinction thresholds in demographic models of territorial populations. *The American Naturalist*. **130**: 624–635.

Lande R., 1988. Demographic models of the Northern Spotted Owl (Strix occidentalis caurina). *Oecologia*. **75**: 601–607.

Levins R., 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*. **15**: 237–240.

Levins R., 1970. Extinction. 75-107 In: *Some mathematical problems in biology* (M. Gerstenhaber, *ed.*). Providence, RI: American Mathematical Society.

MacKenzie D.I. & Nichols J.D., 2004. Occupancy as a surrogate for abundance estimation. *Animal Biodiversity and Conservation*. **27**: 461-467.

MacKenzie D.I., Nichols J.D., Lachman G.B., Droege S., Royle A.J. & Langtimm A.C., 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*. **83**: 2248 – 2255.

MacKenzie D.I., Nichols J.D., Royle J.A., Pollock K.H., Bailey L.L. & Hines J.E., 2006. *Occupancy Estimation and Modeling*. Academic Press. San Diego, California, USA. 324p.

Nichols, J. D., 2004. Evolution of quantitative methods for the study and management of avian populations : on the importance of individual contributions. *Animal Biodiversity and Conservation*. **27(1)**: 3–19.

Okubo A., 1986. Diffusion-type models for avian range expansion. Pages 1038–1049 *in* Henri Ouellet, editor. Acta XIX Congressus Internationalis Ornithologici. National Museum of Natural Sciences, University of Ottawa Press, Ottawa, Ontario, Canada.

Pulliam H.R., 1988. Sources, sinks, and population regulation. *The American Naturalist*. **132**: 652—661.

Royle J.A. & Dorazio R.M., 2006. Hierarchical models of animal abundance and occurrence. *Journal of Agricultural, Biological, and Environmental Statistics*. **11(3)**: 249—263.

Royle J.A. & Dorazio R.M., 2008. *Hierarchical Modeling and inference in ecology (The analysis of data from populations, metapopulations, and communities)*. Academic Press.

Royle J.A. & Kéry M., 2007. A Bayesian state-space formulation of dynamic occupancy models. *Ecology*. **88(7)**: 1813—1823.

Sahu S.K., Gelfand A.E. & Holland D.M., 2006. Spatio-temporal modeling of fine particulate matter. *Journal of Agricultural Biology and Environmental Statistics*. **11**: 61—86.

Sahu S.K., Jona Lasinio G., Orasi A. & Mardia, K.V., 2005. A comparison of spatio-temporal Bayesian models for reconstruction of rainfall fields in a cloud seeding experiment. *Journal of Mathematics and Statistics*. **1(4)**: 273—281.

Sahu S.K. & Mardia K.V., 2005. A Bayesian Kriged-Kalman model for short-term forecasting of air pollution levels. *Journal of the Royal Statistical Society, Series C*. **54**: 223–244.

Sansò B. & Guenni L., 1999. Venezuelan rainfall data analysed by using a Bayesian space-time model. *Applied Statistics*. **48**: 345–362.

Sansò B. & Guenni L., 2000. A nonstationary multisite model for rainfall. *Journal of American Statistics Association*. **95**: 1089-1100.

Sauer J. R., Peterjohn B.G. & Link W.A., 1994. Observer differences in the North American Breeding Bird Survey. *The Auk*. **111(1)**: 50—62.

van Straalen N.M., 1998. Evaluation of bioindicators systems derived from soil arthropod communities. *Applied Soil Ecology*. **9**: 429—437.

Veit R.R. & Lewis M.A., 1996. Dispersal, Population Growth, and the Allee Effect: Dynamics of the House Finch Invasion of Eastern North America. *The American Naturalist*. **148**(2): 255-274.

Weber D., Hintermann U. & Zangger A., 2004. Scale and trends in species richness: considerations for monitoring biological diversity for political purposes. *Global Ecology and Biogeography*. **13**: 97—104.

Wikle C.K., 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*. **84(6)**: 1382—1394.

Wikle C.K., Milliff R.F., Nychka D. & Berliner L.M., 2001. Spatiotemporal hierarchical Bayesian modeling: tropical ocean surface winds. *Journal of the American Statistical Association*. **96**: 382–397.

Williams B.K., Nichols J.D. & Conroy M.J., 2002. *Analysis and Management of Animal Populations*. Academic Press. San Diego, California, USA. 817p.

APPENDICES

APPENDIX

Chapter II

Appendix A.

Appendix a. WinBugs code for the model of invasion process.

```
model {

  ############################################################
  #                                                          #
  #                  Ecological process                      #
  #                                                          #
  ############################################################

 ############ Parameters initialization ########################
psi~dunif(0,1)                  # initial probability of occupancy
lphi1~dnorm(0,.1)               # "rescue effect" survival 1
lphi2~dnorm(0,.1)               # "rescue effect" survival 2
lgamma1~dnorm(0,.1)             # "rescue effect" colonization parameter 1
lgamma2~dnorm(0,.1)             # "rescue effect" colonization parameter 2
ltheta1~dnorm(0,.1)             # "rescue effect" recolonization parameter 1
ltheta2~dnorm(0,.1)             # "rescue effect" recolonization parameter 1

alpha~dunif(0,1)                # Layer 1's density weight
beta~dunif(0,1)                 # Layer 2's density weight

pi2 <- 3.14159*2
dirSpread~dunif(0,pi2)

 for(i in 1:(nyear-1))
 {
     lphi0[i]~dnorm(0,1)          # intrinsic survival
     lgamma0[i]~dnorm(0,1)        # intrinsic colonization probability
     ltheta0[i]~dnorm(0,1)        # intrinsic recolonization probability
 }

############ Model ########################

 for(i in 1:nsite)
 {
     z[i,1]~dbern(psi)            # occupancy status initialization
     A[i,1]<- 1-z[i,1]            # availability (recruitability)

     for(t in 2:nyear)
     {
         ### defining local density###

         #Layer 1
         for (j in 1:N1[i,1])
         {
          ConnSite1[i,j,t-1] <- z[layer1[i,j],t-1]*(1-
cos(dirSpread)*Cosinus[i,layer1[i,j]]-sin(dirSpread)*Sinus[i,layer1[i,j]])
# connected neighbors in layer 1 weighted by its position to spread
direction
         }
         D1[i,t-1]<- sum(ConnSite1[i,1:N1[i,1],t-1])/N1[i,1]     #density
'rate' in first layer

         #Layer2
         for (k in 1:N2[i,1])
```

```
         {
          ConnSite2[i,k,t-1] <- z[layer2[i,k],t-1]*(1-
cos(dirSpread)*Cosinus[i,layer2[i,k]]-sin(dirSpread)*Sinus[i,layer2[i,k]])
# connected neighbors in layer 2 weighted by its position to spread
direction
         }
         D2[i,t-1]<- sum(ConnSite2[i,1:N2[i,1],t-1])/N2[i,1]     #density
'rate' in second layer


         #local density estimator
         D[i,t-1]<- alpha*D1[i,t-1] + beta*D2[i,t-1]



         ### defining parameters ###
          phi[i,t-1]  <- lphi0[t-1]+D[i,t-1]*lphi1+D[i,t-1]*D[i,t-1]*lphi2
          gamma[i,t-1] <- lgamma0[t-1]+D[i,t-1]*lgamma1+D[i,t-1]*D[i,t-
1]*lgamma2
          theta[i,t-1] <- ltheta0[t-1]+D[i,t-1]*ltheta1+D[i,t-1]*D[i,t-
1]*ltheta2



         ### defining occupancy probability  ###
         logit(muZ[i,t])<-   z[i,t-1]*phi[i,t-1]                #  survival
if occupied
                    + (1-z[i,t-1])*gamma[i,t-1]*A[i,t-1]        # first
colonization   (creation if never occupied)
                   + (1-z[i,t-1]-A[i,t-1])*theta[i,t-1]         #
recolonization    (re creation if not occupied before but has already been
occupied)

         z[i,t]~dbern(muZ[i,t])           #    new occupancy status
         A[i,t]<-(1-z[i,t])*A[i,t-1]      #    new availability


     }
 }

  ###############################################################
  #                                                             #
  #                  Observation process                        #
  #                                                             #
  ###############################################################


 ############ Parameters initialization ########################

 ObsInit~dnorm(0,.1)
 DeltaObs~dnorm(0,.1)

 ############ Model ########################


   for(i in 1:nsite)
    {

       Y[i,1]~dbin(prob[i,1],K[i,1])
       prob[i,1]<-p[i,1]*z[i,1]        #p[i,t]=detection probability ;
K[i,t]=number of visits, z[i,t]=occupancy status
```

```
       logit(p[i,1])<-ObsInit          #  Beginning of study based on no
previous detection
       Aprim[i,1]<-1-Y[i,1]/max(1,Y[i,1])   # Aprim[i,t-1]=1 if available
for a first detection at time t, i.e. if no one have ever been detected on
site i before

       for(t in 2:nyear)
       {
       Y[i,t]~dbin(prob[i,t],K[i,t])
       prob[i,t]<-p[i,t]*z[i,t]
       logit(p[i,t])<-ObsInit+DeltaObs*(1-Aprim[i,t-1])

       Aprim[i,t]<-(1-Y[i,t]/max(1,Y[i,1]))*Aprim[i,t-1]      #    new
previous detection status
       }
    }

}
```

APPENDIX

Chapter III

Appendices

Appendix 1. Parameters used and estimated in the selected model.

Parameter	Description	Algorithm notation
$Z_{i,t}$	Occupancy state of cell i during year t.	`z[i,t-1]`
$\mu_{i,t}$	Site occupancy probability	`muZ[i,t]`
$\phi_{i,t}$	Persistence parameter	`phi[i,t-1]`
$\gamma_{i,t}$	First colonization parameter	`gamma[i,t-1]`
$\theta_{i,t}$	Recolonization parameter	`theta[i,t-1]`
$A_{i,t}$	Availability of a site for first colonization	`A[i,t-1]`
$S_{i,t}$	Local success status	`Success[i,t-1]`
$D_{i,t}$	Local density	`D[i,t-1]`
$\tau_{i,t}$	Neighboring success rate	`Srate[i,t-1]`
w_{ij}	Connection matrix	`w[i,j]`
a_t, b_t, c_t	Respectively intrinsic persistence, first colonization and recolonization parameters	`lphi0[t]`, `lgamma0[t]`, `ltheta0[t]`
ϕ_1, γ_1	Respectively linear factors for relationship between persistence, first colonization and local density	`lphi1, lgamma1`
ϕ_2	Quadratic coefficient for persistence/local density relationship	`lphi2`
ϕ_S, γ_S, θ_S	Relationship between neighboring success and respectively persistence, first colonization and recolonization	`lphiNS`, `lgammaNS`, `lthetaNS`
$\phi_{\text{int}2}$	Interaction between local and neighboring success for persistence parameter	`lphiinter2`

Appendix 2. WinBugs code for the selected model of nesting site occupancy for the kittiwake.

(*input:* $Z_{i,t}$, $S_{i,t}$ *and* w_{ij})

```
model {

############ Parameters initialization ########################
psi~dunif(0,1)                  # initial probability of occupancy

lphi1~dnorm(0,.1)               # linear relationship between persistence
and density
lphi2~dnorm(0,.1)               # quadratic relationship between persistence
and density
lgamma1~dnorm(0,.1)             # linear relationship between first
colonization and density

lphiNS~dnorm(0,.1)              # impact of neighboring success on
persistence
lgammaNS~dnorm(0,.1)            # impact of neighboring success on first
colonization
lthetaNS~dnorm(0,.1)            # impact of neighboring success on
recolonization

lphiinter2~dnorm(0,.1)           # interaction between density and
neighboring success for persistence

for(i in 1:(nyear-1))
   {
   lphi0S[i]~dnorm(0,1)           # intrinsic persistence for local breeding
success
   lphi0F[i]~dnorm(0,1)           # intrinsic persistence for local breeding
success
   lgamma0[i]~dnorm(0,1)          # intrinsic first colonization
   ltheta0[i]~dnorm(0,1)          # intrinsic recolonization
   }

############ Model ########################
for(i in 1:nsite)
   {
   z[i,1]~dbern(psi)             # occupancy status initialization
   A[i,1]<- 1-z[i,1]             # availability (recruitability)

   for(t in 2:nyear)
      {
         ### defining local density ###
      for(j in 1:nsite)
         {
         ConnSite[i,j,t-1] <- z[j,t-1]*w[i,j]   # occupancy status of site j
weighted by connection matrix
         }

      D1[i,t-1]<- sum(ConnSite[i,1:nsite,t-1])
      D[i,t-1]<-(D1[i,t-1])/10                  # scaled local density

         ### defining neighboring success rate ###
      for(j in 1:nsite)
         {
         NeighborSuccess[i,j,t-1] <- Success[j,t-1]*w[i,j]      # succes
status of site weighted by connection matrix
         }
```

```
1
2       Srate[i,t-1]<-sum(NeighborSuccess[i,1:nsite,t-
3   1])/max(1,sum(ConnSite[i,1:nsite,t-1]))   # neighboring success rate
4
5          ### defining parameters ###
6       #persistence parameter
7       logit(phi[i,t-1])  <- lphi0S[t-1]*Success[i,t-1]+lphi0F[t-1]*(1-
8   Success[i,t-1])
9                             +D[i,t-1]*lphi1  +D[i,t-1]*D[i,t-1]*lphi2
10                            +Srate[i,t-1]*lphiNS
11                            +lphiinter2*Success[i,t-1]*Srate[i,t-1]
12
13      #first colonization parameter
14      logit(gamma[i,t-1]) <- lgamma0[t-1]
15                              +D[i,t-1]*lgamma1
16                              +Srate[i,t-1]*lgammaNS
17
18      #recolonization parameter
19      logit(theta[i,t-1]) <- ltheta0[t-1]
20                              +Srate[i,t-1]*lthetaNS
21
22         ### defining occupancy probability ###
23      muZ[i,t]<- z[i,t-1]*phi[i,t-1]                      #  persistence event
24                 + (1-z[i,t-1])*gamma[i,t-1]*A[i,t-1]    # first
25  colonization event
26                 + (1-z[i,t-1]-A[i,t-1])*theta[i,t-1]    # recolonization
27  event
28
29      z[i,t]~dbern(muZ[i,t])           #    new occupancy status
30      A[i,t]<-(1-z[i,t])*A[i,t-1]      #    new availability
31
32      }
33    }
34  }
```

APPENDIX

Chapter IV

1

1. Web Appendix A

For convenience, the conditioning on φ_s is omitted in the calculations. Recall that one works conditionally on the presence of species s in R, consequently $\xi_s=1$, or equivalently $\mathbf{z_s} \neq \vec{0}$. Results 1.1 and 1.3 below are mainly a consequence of that:

$$p(\boldsymbol{a_s}) = \frac{\varphi_s^{A_s}(1-\varphi_s)^{J_a-A_s}}{1-(1-\varphi_s)^{J}} \quad (1)$$

where $\boldsymbol{a_s}$ denotes a vector of length J_a extracted from $\mathbf{z_s}$, and different from the null vector (thus it includes at least one component equal to 1), and A_s denotes the sum of all the components of $\boldsymbol{a_s}$. The above result is easily obtained by suming $p(\boldsymbol{a_s}, \boldsymbol{b_s})$ over all the possible values of $\boldsymbol{b_s}$, where $\boldsymbol{b_s}$ denotes a vector extracted from $\mathbf{z_s}$, such that $\boldsymbol{b_s}$ and $\boldsymbol{a_s}$ constitute a partition of $\mathbf{z_s}$. Note that (1) cannot be used if $\boldsymbol{a_s} = \vec{0}$ since, in such a case, the sum has to be done over all the possible values of $\boldsymbol{b_s}$ different from the null vector (considering that $\mathbf{z_s} \neq \vec{0}$).

1.1 $\Pr(z_{sj} = 1) = \varphi_s/[1-(1-\varphi_s)^J]$ is deduced from (1) by assuming that $\boldsymbol{a_s}$ is reduced to the j-th component of $\mathbf{z_s}$ and that $z_{sj} = 1$.

1.2 From now, it is convenient to denote random variables by capital letters. We introduce the vector $\overline{z_{sj}}$ of length $J-1$ and equal to the vector $\mathbf{z_s}$ from which the j-th component has been removed. We have:

$$\Pr\left(Z_{sj} = 1 \middle| \overline{\boldsymbol{Z}_{sj}} \neq \vec{0}\right) = \frac{\Pr(Z_{sj} = 1 \text{ and } \overline{\boldsymbol{Z}_{sj}} \neq \vec{0})}{\Pr(\overline{\boldsymbol{Z}_{sj}} \neq \vec{0})}$$

By observing that $\Pr(\overline{\boldsymbol{Z}_{sj}} \neq \vec{0}) = \Pr(\boldsymbol{Z_s} = \boldsymbol{z_s})$ where $\mathbf{z_s}$ is such that $z_{si} = 0$ for all $i \neq j$ and $z_{sj} = 1$ (since $\mathbf{z_s} \neq \vec{0}$), and by using (3.3), one has:

$$\Pr(\overline{\boldsymbol{Z}_{sj}} \neq \vec{0}) = 1 - \frac{\varphi_s\ (1-\varphi_s)^{J-1}}{1-(1-\varphi_s)^{J}} = \frac{1-(1-\varphi_s)^{J-1}}{1-(1-\varphi_s)^{J}}$$

Recall that if A and B denote two any events of a probability space, one has $\Pr(A \cap B) = 1-\Pr(\overline{A \cap B})$ and $\Pr(A \cap B) = \Pr(\bar{A} \cup \bar{B}) = \Pr(\bar{A}) + \Pr(\bar{B}) - \Pr(\bar{A} \cap \bar{B})$. Therefore we have:

$$\Pr(Z_{sj} = 1 \text{ and } \overline{\mathbf{Z}_{sJ}} \neq \vec{0}) = 1 - [\Pr(Z_{sj} = 0) + \Pr(\overline{\mathbf{Z}_{sJ}} = \vec{0})],$$

by observing that $\Pr(Z_{sj} = 0 \text{ and } \overline{\mathbf{Z}_{sJ}} = \vec{0}) = 0$ since $\mathbf{z_s} \neq \vec{0}$. By using now the expressions of $\Pr(\overline{\mathbf{Z}_{sJ}} = \vec{0})$ and $\Pr(Z_{sj} = 1)$, we deduce that:

$$\Pr(Z_{sj} = 1 \text{ and } \overline{\mathbf{Z}_{sJ}} \neq \vec{0}) = \varphi_s \frac{1 - (1 - \varphi_s)^{J-1}}{1 - (1 - \varphi_s)^{J}}$$

hence the result.

1.3 We now prove the Proposition 3.1. Let *i* and *j* be two distinct quadrats. We denote by *E* the following event: *it exits at least one quadrat k distinct from i and j such that* $z_{sk} = 1$. To prove that z_{si} and z_{sj} are independent, conditionally on *E*, we have to show that $p(z_{si}, z_{sj} \mid E) = p(z_{sj} \mid E)\, p(z_{sj} \mid E)$ for all pairs (z_{si}, z_{sj}). Taking into account (1), it is easy to check that:

$$\Pr(E) = \frac{1 - (1 - \varphi_s)^{J-2}}{1 - (1 - \varphi_s)^{J}}$$

and that

$$\sum_{v} \Pr(Z_{si} = z_{si}, Z_{sj} = z_{sj}, V = v) = \varphi_s^{z_{si}+z_{sj}} (1 - \varphi_s)^{2 - z_{si} - z_{sj}} \Pr(E)$$

where *v* denotes any vector composed of 0 and 1, distinct from $\vec{0}$, and having $J - 2$ components; from which we deduce that $p(z_{si}, z_{sj} \mid E) = \varphi_s^{z_{si}+z_{sj}} (1 - \varphi_s)^{2 - z_{si} - z_{sj}}$, To calculate $p(z_{si} \mid E)$ and $p(z_{sj} \mid E)$ we proceed as for $p(z_{si}, z_{sj} \mid E)$ and so we have: $p(z_{si} \mid E) = \varphi_s^{z_{si}} (1 - \varphi_s)^{1 - z_{si}}$ and $p(z_{sj} \mid E) = \varphi_s^{z_{sj}} (1 - \varphi_s)^{1 - z_{sj}}$; hence the result.

2. Web Appendix B

The aim of this brief study is to show that the probability that a species present in R is not detected (or very few detected) can be relatively high when it is spatially rare or hard to detect. In this paper a target species s is said to be spatially rare if N_s is small, that is if it is present in a small number of quadrats. We have considered that it is the case if $\varphi_s \in [0.05, 0.2]$, that is to say $2 \leq J\ \varphi_s \leq 4$, according to the value of J (see Table 1). Of course, such intervals are somewhat arbitrary (the notion of spatially rare species being actually quite relative); however, such values of φ_s are not unusual in some animal populations, such as birds; see for example: the analysis of the avian point count data in MacKenzie et al. (2006); Dupuis and Joachim (2006); or our illustration, in Section 5. Throughout this study K is fixed ($K = 4$, as in our illustration); moreover, in a first step, the ratio T/J and q_s are also fixed: $T/J = 1$ and $q_s = 0.1$. This corresponds to species relatively difficult to detect at the visit scale; but the relevant quantity here is not q_s, but $\mu_s = 1 - (1 - q_s)^K$; now, in our case $\mu_s = 0.34$, which corresponds to relatively discrete species at the quadrat scale. Table 1 provides, for different values of J, the probabilities of pairs $(V_s,W_s) = (0, 0)$ and $(V_s,W_s) = (1, 1)$.

We effectively observe that the probabilities to observe the pairs (0, 0) and (1, 1) are relatively important (even for relatively large values of $_J$). If one considers spatially rare species, but easily detectable, the probabilities that it is detected during one visit only (even not detected) are not necessarily small. For example: $p(0, 0) = 0.20$ and $p(1, 1) = 0.29$, if $J = T = 20$, $\varphi_s = 0.1$, $q_s = 0.2$ (that is $\mu_s \approx 0.60$); and $p(0, 0) = 0.10$ and $p(1, 1) = 0.21$, if $J = T = 40$, $\varphi_s = 0.05$, $q_s = 0.3$ (that is $\mu_s \approx 0.75$). In Table 1, the fraction T/J of sampled quadrat is fixed (since $T = J$); it is clear that if we reduce this fraction we will increase the above probabilities. For example, $p(0, 0) = 0.38$ and $p(1, 1) = 0.23$, if $J = 40$ and $T/J = 0.50$ (the values of q_s and φ_s being those of the last example).

3. Web Appendix C

Proof of the proposition 4.1.

Recall that a statistical model M parametrized by θ is identifiable if

$$p(.|\theta_1) = p(.|\theta_2) \Rightarrow \theta_1 = \theta_2$$

where the hypothesis $p(.|\theta_1) = p(.|\theta_2)$ means that: $p(y|\theta_1) = p(y|\theta_2)$ whatever the data set **y** (e.g. Lehman, 1989; page 456). It is equivalent to say that θ is identifiable. Consequently, to show that q_s and φ_s are both identifiable, we have to prove that:

$$[\ \forall\, y_s \in Y,\ p(y_s|\ \varphi_s = \varphi_1,\ q_s = q_1) = p(y_s|\ \varphi_s = \varphi_2,\ q_s = q_2] \Rightarrow \varphi_1 = \varphi_2 \text{ and } q_1 = q_2,$$

where Y denotes the set in which y_s takes its values, where q_1 and q_2 designate two any values of q_s, and φ_1 , φ_2 designate two any values of φ_s. To prove identifiability it is necessary to distinguish two cases: $K = 1$ and $K \geq 2$.

First case: $K = 1$. We consider two particular records. The first one, called $\mathbf{y_s}$, is such that species s has been detected in the T sampled quadrats (therefore $y_{sj} = 1$ for all j). The second one, called $\mathbf{y'_s}$, is such that species s has been detected in the $T - 1$ quadrats (therefore $y_{sj} = 1$ for all j, except one). Using (4.1), one has:

$$\Pr(\mathbf{y}_s|\varphi_1, q_1\ ) = \frac{(\varphi_1 q_1)^T}{1-(1-\varphi_1)^J}$$

and

$$\Pr(\mathbf{y'}_s|\varphi_1, q_1\ ) = \frac{(\varphi_1 q_1)^{T-1}(1-\varphi_1 q_1)}{1-(1-\varphi_1)^J}$$

Now, by hypothesis, $p(y_s|\varphi_1, q_1) = p(y_s|\varphi_2, q_2)$ and $p(y'_s|\varphi_1, q_1) = p(y'_s|\varphi_2, q_2)$; hence

$$\frac{p(y_s|\varphi_1, q_1)}{p(y'_s|\varphi_1, q_1)} = \frac{p(y_s|\varphi_2, q_2)}{p(y'_s|\varphi_2, q_2)}$$

After simplification, we obtain:

$$\varphi_1 q_1 = \varphi_2 q_2. \quad (2)$$

Using again the hypothesis $p(y_s|\varphi_1, q_1) = p(y_s|\varphi_2, q_2)$ and the equality (2), we deduce that: $1-(1-\varphi_1)^J = 1-(1-\varphi_2)^J$, hence $\varphi_1 = \varphi_2$. Using again (2), we deduce that $q_1 = q_2$. Consequently, q_s and φ_s are both identifiable.

Second case: $K \geq 2$. *(This part of the proof is just outlined, for concision.)*

Parameter q_s is proved to be identifiable, by considering the two following records. The first one (called *h1*) is such that species *s* has been detected at each visit; in other terms, $y_{sj} = K$ for all *j*. The second one (called *h2*) is such that species *s* has been detected at each visit apart from once (in other terms, $y_{sj} = K$ for all quadrat *j*, except one, say *i*, for which $y_{si} = K - 1$.

Parameter φ_s is proved to be identifiable, by considering the two following records: the first one is *h1*, and the second one, called *h3*, is such that $y_{sj} = K$ for all *j*, except in one, say *i*, for which $y_{si} = 0$.

4. Web Appendix D

Proof of the proposition 4.2.

This part of the proof is again just outlined (for concision). It uses the records defined in Appendix C. The proof includes three steps. First q_s is proved to be identifiable, then φ_s, and finally λ_s. To prove that q_s is identifiable use *h1* and *h2*, for φ_s use *h1* and *h3* (all defined in Appendix C). Then, we immediately deduce that λ_s is identifiable by using *h1* (for example), and the fact that q_s and φ_s are identifiable.

5. Web Appendix E

The augmentation data step for the conditional approach.

Implementing this step requires a particular attention and is thus described in details. For each species $s \in S$, it consists in simulating z_s^{mis} according to $p(z_s^{mis} | y_s, q_s, \varphi_s)$. Two cases have to be distinguished depending on whether $y_s = \vec{0}$ or not.

First case: $\mathbf{y_s} \neq \vec{0}$ (species *s* has thus been detected in at least one sampled quadrat).

First note that the missing z_{sj} 's are independent conditionally on $\mathbf{y_s}$. It is an immediate consequence of the Proposition 3.1, and of that $\mathbf{y_s} = \vec{0}$ implies that it exits at least on sampled quadrat in which species *s* is present. Thus we can simulate the missing data z_s^{mis} by block, as follows. The first block, which concerns the missing z_{sj} 's where *j* denotes a sampled quadrat, is simulated according to a Binomial $(T - U_s, \omega_s)$, where $\omega_s = \Pr(z_{sj} = 1 | x_{sj}, z_s^{obs}, q_s, \varphi_s)$. By applying the Bayes theorem and by using the fact that $\Pr(z_{sj} = 1 | z_s^{obs}, \varphi_s) = \varphi_s$ (which is a direct consequence of the Proposition 3.1) we deduce that:

$$\omega_s = \frac{\varphi_s(1 - q_s)^K}{\varphi_s(1 - q_s)^K + (1 - \varphi_s)}$$

The other block, which concerns the missing z_{sj} 's where *j* denotes an unsampled quadrat, is simulated according to a Binomial $(J - T, \varphi_s)$.

Second case: $\mathbf{y_s} = \vec{0}$.

Contrary to the previous case, the missing z_{sj} 's are no more independent (conditionally on $\mathbf{y_s}$). Note also that when $\mathbf{y_s} = \vec{0}$, all the z_{sj} 's are actually missing and that we have to simulate the whole vector $\mathbf{z_s}$. This relies on the distribution of $\mathbf{z_s}|\mathbf{y_s} = \vec{0}$, below.

$$\Pr(\mathbf{z_s}|\mathbf{y_s} = \vec{0}) = \frac{\rho_s(1 - q_s)^{Kn_s}\varphi_s^{N_s}(1 - \varphi_s)^{JN_s}}{((1 - q_s)^K\varphi_s + 1 - q_s)^T - (1 - \varphi_s)^J}$$

This expression is deduced from (4.6) in which we have made $W_s = 0$, and from (4.2).

6. Web Appendix F

The MCMC algorithm for the unconditional approach.

When one works unconditionally, the MCMC algorithm implemented for the conditional approach has to be modified. Updating the parameters q_s, φ_s and λ_s presents no difficulty, and details are omitted. The data augmentation step proceeds as follows. As previously, two cases have to be considered, depending on whether $\mathbf{y}_s = \vec{0}$ or not. When $\mathbf{y}_s \neq \vec{0}$ the data augmentation step is not modified. When $\mathbf{y}_s = \vec{0}$ the indicator ξ_s is missing, and missing data now includes both $\mathbf{z}_s$ and ξ_s. The data augmentation step proceeds as follows:

$$\xi_s \sim \xi_s|\mathbf{y}_s, \varphi_s, q_s, \lambda_s \text{ and } \mathbf{z}_s|\mathbf{y}_s, \xi_s, \varphi_s, q_s.$$

By applying the Bayes formula, one has:

$$\Pr(\xi_s = 1 \,|\, \mathbf{y}_s = \vec{0}, \varphi_s, q_s, \lambda_s) = \frac{\lambda_s \Pr(\mathbf{y}_s = \vec{0} \,|\, \xi_s = 1, \varphi_s, q_s)}{\lambda_s \Pr(\mathbf{y}_s = \vec{0} \,|\, \xi_s = 1, \varphi_s, q_s) + (1 - \lambda_s)}$$

where $\Pr(\mathbf{y}_s = \vec{0} \,|\, \xi_s = 1, \varphi_s, q_s)$ is given by (4.2). The distribution of $\mathbf{z}_s|\mathbf{y}_s, \xi_s, \varphi_s, q_s$ is given by (1) in Appendix E, when $\xi_s = 1$. If $\xi_s = 0$, thus $\mathbf{z}_s = \vec{0}$ (with probability one).

References

Lehmann, E.L. (1999). Elements of Large-sample Theory. Springer (New York).

Table 1

Probabilities that a species present in the region R is not detected (i.e. v = w = 0) or is detected during only one visit (i.e. v = w = 1), in function of J and φ

J	*φ*	*(v,w)*	*P(v,w)*
10	0.2	(0,0)	0.43
10	0.2	(1,1)	0.34
20	0.2	(0,0)	0.23
20	0.2	(1,1)	0.30
40	0.1	(0,0)	0.23
40	0.1	(1,1)	0.30
80	0.05	(0,0)	0.17
80	0.05	(1,1)	0.26

APPENDIX

Chapter V

1

Appendix 1. MatLab code for the binary search algorithm to determine coefficient of the Beta distribution for the a priori mean and 95% confidence interval on detection probability over the whole visit μ_{sa}. *Output: Beta distribution parameters a and b.*

```
function [ParamBetaMu] = BinarySearchAlgorithm()

mu=input('A priori mean: ')
CIinf=input('Lower bound of a priori CI95%: ')
CIsup=input('Upper bound of a priori CI95%:')
prec=input('Required interval precision: ')

CI=[CIinf CIsup]  ;


% Initialization
margin=prec+1 ;
lambda1=0    ;
lambda2=10000 ;

% Binary search algorithm

while (margin>prec)
      lambda=(lambda1+lambda2)/2     ;
      a=lambda*mu                ;
      b=lambda*(1-mu)           ;
      prob=betainc(CIsup,a,b)-betainc(CIinf,a,b)     ;
      margin=abs(proba-0.95)                ;

      if (prob<0.95)
         lambda1=lambda             ;
   end

   if (prob>0.95)
      lambda2=lambda        ;
   end
end

a=lambda*mu           ;
b=lambda*(1-mu)       ;
ParamBetaMu=[a b]           ;
```

Appendix 2. MatLab code to obtain the Beta distribution parameters for the a priori on q_{sa} (detection probability during each session of a visit) from the Beta distribution parameters for the a priori on μ_{sa}. ***Input:*** *Beta distribution parameters on μ_{sa} a priori provided by function BinarySearchAlgorithm().* ***Output:*** *Beta distribution parameters for q_{sa} a priori.*

```
function [ParamBetaQ]=MuGlobal_QLocal(ParamBetaMu)

a= ParamBetaMu (1);
b= ParamBetaMu (2);

K=input('enter the number of sessions K during a visit:');
L=input('enter the number of iterations for the MCMC algorithm:');

%Initialization
sumq=0;
sum2q=0;

%Beginning of MCMC algorithm

for l=1:L
   X=betarnd(a,b);
   k=1/K;
   q=1-(1-X)^k;
   Y(l)=q;
   sumq=sumq+q;
   sum2q=sum2q+q*q;
   espq(l)=sumq/l;
end;

%End of MCMC algorithm

meanq=sumq/L;
varq=sqrt(sum2q/L-meanq^2);
lambq=((meanq*(1-meanq))/varq^2)-1;

alpha=lambq*meanq;
beta=lambq*(1-meanq);

Force=alpha+beta;
APrioriMean=alpha/(alpha+beta) ;

ParamBetaQ =[APrioriMean  Force];
```

Appendix 3. MatLab code for estimating γ_{sa} (and φ_{sa} and q_{sa} if required) using the MCMC methods, Hasting-Metropolis algorithm for updating φ_{sa} , and Gibbs algorithm for updating q_{sa}. ***Input:*** *Total number Vobs of quadrats where the species of interest has been detected, total number W of detection of targeted species, total number J of quadrats in region R_a, number T of sampled quadrats, number K of sessions per visit, number L of iterations for the MCMC algorithm.* ***Output:*** *Estimates of occupancy rate γ_{sa} and a posteriori confidence interval.*

```
function
[estimgamma,ICgamma1,ICgamma2,ICgamma3]=DBJDataAugV1V2(Vobs,W,J,T,K,L)

delta=input('enter value of adjustement parameter on phi for Hasting-
Metropolis algorithm:');

% Computation of Beta distribution parameter of a priori on Qsa
[parammu]= BinarySearchAlgorithm () ;
[paramq]= MuGlobal_QLocal(parammu)  ;
priorq=paramq(1);
lambdaq =paramq(2);

%DATA
dataVW=[Vobs,W];
dataVW;

%Non-informative a priori on phi
priorphi=0.5;
lambdaphi=2;

%Initialization
phi=0.1;
phi_init=phi;
phi_init;

q=0.3;
q_init=q;
q_init;

accept=0;
sumq=0;
sum2q=0;
sumphi=0;
sum2phi=0;
sumgamma=0;









%Beginning of MCMC algorithm

for l=1:L


 % 1) DATA AUGMENTATION
 omega=(phi*(1-q)^K)/(phi*(1-q)^K+(1-phi));

 if (Vobs==0);
   Vsim=0;
   while(Vsim==0)
     Vsim1=binornd(T,omega);
```

```
     if (J>T)
       Vsim2=binornd(J-T,phi);
     else
       Vsim2=0;
     end;
     Vsim=Vsim1+Vsim2;
   end;
 else
   if (T>Vobs)
       Vsim1=binornd(T-Vobs,omega);
   else
       Vsim1=0;
   end;
   if (J>T)
      Vsim2=binornd(J-T,phi);
   else
      Vsim2=0;
   end;
   Vsim=Vsim1+Vsim2;
 end;
  V=Vobs+Vsim;

 gamma=V/J;
 grafgamma(l)=gamma;

 % 2) UPDATE $q$ via Gibbs
   q=betarnd(lambdaq*priorq+W,lambdaq*(1-priorq)+(K*Vobs-W)+K*Vsim1);

 % 3) UPDATE $phi$ via Hast. Metro.
e=max(0,phi-delta);
f=min(1,phi+delta);
phihm=(f-e)*rand(1)+e;
ehm=max(0,phihm-delta);
fhm=min(1,phihm+delta);
logratiog=log(fhm-ehm)-log(f-e);

piphihm=betapdf(phihm,priorphi*lambdaphi,(1-priorphi)*lambdaphi);
piphi=betapdf(phi,priorphi*lambdaphi,(1-priorphi)*lambdaphi);
logratioprior=log(piphihm)-log(piphi);

logPrc=V*(log(phihm)-log(phi));
logPrd=(J-V)*(log(1-phihm)-log(1-phi));
logPre=log(1-(1-phi)^J)-log(1-(1-phihm)^J);
logratiolike=logPrc+logPrd+logPre;

logrho=logratiolike+logratioprior-logratiog;
rho=exp(logrho);


if (rho>=1)
    phi=phihm;
    accept=accept+1;
else
    U=binornd(1,rho);
    if(U==1)
      phi=phihm;
      accept=accept+1;
    end;
 end;

sumgamma=sumgamma+gamma;
```

```
sumphi=sumphi+phi;
sum2phi=sum2phi+phi*phi;
sumq=sumq+q;
sum2q=sum2q+q*q;

end;
%End of MCMC algorithm

rate=accept/L;
rate;

estimgamma=sumgamma/L;

estimq=sumq/L;
errorq=sqrt(sum2q/L-estimq^2);

estimphi=sumphi/L;
errorphi=sqrt(sum2phi/L-estimphi^2);

estimationqphi=[estimq,errorq,estimphi,errorphi];

icgamma=sort(grafgamma);

ICgamma1=[icgamma(0.025*L),icgamma(0.975*L)];
ICgamma1;

ICgamma2=[icgamma(0.05*L),icgamma(L)];
ICgamma2;

ICgamma3=[icgamma(1),icgamma(0.95*L)];
ICgamma3;

estimgamma ;
```

Title:
Tests of hypotheses in fragmentated population dynamics:
development and applications of site occupancy models

Abstract:

Classical approaches to the development of spatial models for binary processes of species distribution (i.e. occupancy processes) present three important deficiencies. i) They do not explicitly accommodate sampling uncertainty. ii) There is a lack of spatio-temporal occupancy models, especially in the framework of hierarchical modeling. iii) Most of existing models are phenomenological and do not explicitly consider underlying ecological mechanisms.

This thesis develops spatio-temporal occupancy models for dynamical ecological processes in order to respond to these limitations while incorporating scientific knowledge in every modeling step. Those models are applied to critical ecological topics ranging from the spread of invasive species to habitat selection via climate changes.

Understanding range and occupancy dynamics will permit prediction of occupancy changes that are likely to accompany future changes and hopefully will permit informed attempts to mediate changes in occupancy.

AUTEUR : Florent BLED

TITRE : Tests d'hypothèses en dynamique des populations fragmentées: développement et applications de modèles d'occupation des sites

DIRECTEURS DE THESE : Emmanuelle Cam et Jérôme Dupuis

LIEU ET DATE DE SOUTENANCE : Toulouse, le 18 mai 2010

RESUME

Les approches classiques de modèles spatiaux pour les processus binaires de distribution d'espèces (i.e. occupation des sites) présentent trois importantes carences. i) Elles ne prennent pas explicitement en compte l'incertitude dans le processus d'échantillonnage. ii) Il y a un manque de modèles spatio-temporels, notamment hiérarchique. iii) La plupart des modèles existants sont de type phénoménologique et ne considèrent pas explicitement les mécanismes écologiques sous-jacents.

Cette thèse répond à ces limitations en présentant des modèles spatio-temporels d'occupation des sites pour des processus écologiques dynamiques. Ces modèles sont appliqués à des sujets essentiels en écologie, tels que la sélection de l'habitat, les espèces invasives et les changements climatiques.

Comprendre la dynamique d'occupation des sites permet de prédire les changements d'occupation qui accompagneront des modifications de l'habitat et de prendre des décisions adaptées en gestion des populations.

MOTS-CLES

Occupation des sites, Détectabilité, Modèle hiérarchique, Statistiques bayésiennes, Espèce invasive, Sélection de l'habitat, Changement climatique, Dynamique des populations

---

DISCIPLINE ADMINISTRATIVE : Ecologie, Biodiversité, Evolution

---

INTITULE ET ADRESSE DE L'U.F.R. OU DU LABORATOIRE :

UMR 5174 Evolution et Diversité Biologique
Bâtiment 4R3 B2
Université Paul Sabatier
118, route de Narbonne
31062 Toulouse cedex 9, France